COMPUTED TOMOGRAPHY

Is it possible to use low-dose deep learning reconstruction for the detection of liver metastases on CT routinely?

Peijie Lyu[1,2] · Nana Liu[1] · Brian Harrawood[3] · Justin Solomon[3] · Huixia Wang[1] · Yan Chen[1] · Francesca Rigiroli[4] · Yuqin Ding[2,5] · Fides Regina Schwartz[2] · Hanyu Jiang[2,6] · Carolyn Lowry[7] · Luotong Wang[8] · Ehsan Samei[3] · Jianbo Gao[1] · Daniele Marin[2]

Abstract

Objectives To compare the image quality and hepatic metastasis detection of low-dose deep learning image reconstruction (DLIR) with full-dose filtered back projection (FBP)/iterative reconstruction (IR).

Methods A contrast-detail phantom consisting of low-contrast objects was scanned at five CT dose index levels (10, 6, 3, 2, and 1 mGy). A total of 154 participants with 305 hepatic lesions who underwent abdominal CT were enrolled in a prospective non-inferiority trial with a three-arm design based on phantom results. Data sets with full dosage (13.6 mGy) and low dosages (9.5, 6.8, or 4.1 mGy) were acquired from two consecutive portal venous acquisitions, respectively. All images were reconstructed with FBP (reference), IR (control), and DLIR (test). Eleven readers evaluated phantom data sets for object detectability using a two-alternative forced-choice approach. Non-inferiority analyses were performed to interpret the differences in image quality and metastasis detection of low-dose DLIR relative to full-dose FBP/IR.

Results The phantom experiment showed the dose reduction potential from DLIR was up to 57% based on the reference FBP dose index. Radiation decreases of 30% and 50% resulted in non-inferior image quality and hepatic metastasis detection with DLIR compared to full-dose FBP/IR. Radiation reduction of 70% by DLIR performed inferiorly in detecting small metastases (< 1 cm) compared to full-dose FBP (difference: −0.112; 95% confidence interval [CI]: −0.178 to 0.047) and full-dose IR (difference: −0.123; 95% CI: −0.182 to 0.053) ($p < 0.001$).

Conclusion DLIR enables a 50% dose reduction for detecting low-contrast hepatic metastases while maintaining comparable image quality to full-dose FBP and IR.

Key Points

• *Non-inferiority study showed that deep learning image reconstruction (DLIR) can reduce the dose to oncological patients with low-contrast lesions without compromising the diagnostic information.*

• *Radiation dose levels for DLIR can be reduced to 50% of full-dose FBP and IR for detecting low-contrast hepatic metastases, while maintaining comparable image quality.*

• *The reduction of radiation by 70% by DLIR is clinically acceptable but insufficient for detecting small low-contrast hepatic metastases (< 1 cm).*

✉ Jianbo Gao
gjbfsk@126.com

1. Department of Radiology, The First Affiliated Hospital of Zhengzhou University, No.1 East Jianshe Road, Zhengzhou 450052, Henan Province, China

2. Department of Radiology, Duke University Medical Center, 2301 Erwin Rd, Box 3808, Durham, NC 27710, USA

3. Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, 2424 Erwin Road, Suite 302, Durham, NC 27705, USA

4. Beth Israel Deaconess Medical Center Department of Radiology, Harvard Medical School, 1 Deaconess Rd, 330 Brookline Ave, Boston, MA 02215, USA

5. Department of Radiology, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Shanghai 20032, China

6. Department of Radiology, West China Hospital of Sichuan University, 37 Guoxue Alley, Chengdu 610041, Sichuan Province, China

7. Clinical Imaging Physics Group, Duke University Health System, 2424 Erwin Rd, Ste. 302, Durham, NC 27705, USA

8. CT Imaging Research Center, GE Healthcare China, No.1 Tongji South Road, Beijing 100176, China

Abbreviations

2AFC	Two-alternative forced choice
CI	Confidence interval
CNR	Contrast-to-noise ratio
CT	Computed tomography
$CTDI_{vol}$	Volumetric CT dose index
DLIR	Deep learning image reconstruction
FBP	Filtered back projection
FOM	Figure of merit
GEE	Generalized estimating equations
IQR	Interquartile range
IR	Iterative reconstruction
JAFROC	Jackknife alternative free-response receiver operating characteristic

Introduction

Detection of hepatic metastases in computed tomography (CT) is challenging as, usually, there is a small CT number difference between a metastatic lesion and its surroundings, which is described as a "low-contrast detection task." Lower dose abdominal CT images may present additional difficulties due to the increased image noise and the less-defined borders for small low-contrast hepatic lesions. Iterative reconstruction (IR) improves the image quality of low-dose CT images compared to filtered back projection (FBP) reconstruction [1, 2]. However, its nonlinear image denoising and contrast-dependent spatial resolution characteristics degrade the noise texture and low-contrast detectability at low-dose levels [3–5]. This limits observer performance for low-contrast detection tasks (such as detecting low-contrast hypoattenuating liver metastases) [2, 6, 7].

Commercially available artificial intelligence–driven methods have been recently developed to overcome the limitations of the FBP and IR approaches, such as deep learning image reconstruction (DLIR) (True Fidelity, GE Healthcare). The DLIR algorithm utilizes convolutional neural networks, which comprise millions of parameters and an extensive learning process based on the high radiation dose FBP data. On the one hand, DLIR can discriminate true attenuation from noise, improves spatial resolution, and preserves preferable noise texture in CT images [3, 8, 9]. On the other hand, DLIR can generate high-quality images from low-dose projection data in a short reconstruction time in a clinical environment [10, 11].

With the DLIR technique, abdominal CT can be performed under various clinical indications with significant dose reductions while still ensuring diagnostic image quality and

detection of focal lesions [12–15]. Less clear is, however, the ability to correctly detect subtle lesions in seemingly high-quality DLR images that are acquired at a reduced radiation dose. Previous DLIR studies about low-contrast detection were mostly based on phantoms [3, 9, 10]. Jensen et al [14] recently concluded that DLIR improved the CT image quality at 65% radiation dose reduction compared to standard-dose FBP and preserved the ability to detect low-contrast liver lesions larger than 0.5 cm. However, their study was limited to one low-dose level and did not examine correlations of this new algorithm with standard reconstructions across multiple exposures. As there is still a limited amount of data that systematically evaluates the effects of DLIR on the diagnostic performance of low-contrast tasks, it is highly desirable to carry out studies to rigorously determine how much dose reduction can be achieved without degrading diagnostic quality for the given clinical condition.

Therefore, this study began with a multireader comparison of low-contrast hypoattenuating object detection in DLIR to standard reconstructions at various radiation dose levels in a contrast-detail phantom to determine how much radiation it could save. Then, in a larger number of participants, further investigation and validation of DLIR's potential dose reduction was obtained after assessment of image quality and detection of low-contrast hepatic metastases with varying radiation dose levels.

Materials and methods

Phantom experiment

A contrast-detail phantom was used to assess the detectability of low-contrast objects [16] (Fig. 1 A and B, Appendix E1). The phantom was imaged with a 256-row multidetector CT platform (Revolution CT, GE Healthcare) at five dose index levels (10, 6, 3, 2, and 1 mGy, 32 cm CT dose index [CTDI] phantom) (Table 1). Given the phantom size, the dose index levels were selected to obtain phantom images with noise levels roughly equivalent to those of the full-dose clinical images [16]. Images were reconstructed with a 2.5-mm section thickness/interval using FBP, IR (hybrid model-based adaptive statistical iterative reconstruction [ASiR-V] 60%, blended with 40% FBP), and DLIR (medium strength level) [17]

A web-based observer-based image evaluation interface was designed to perform a two-alternative forced-choice (2AFC) detection experiment. Eleven readers (five radiologists and six medical physicists), with 5 to 16 years of
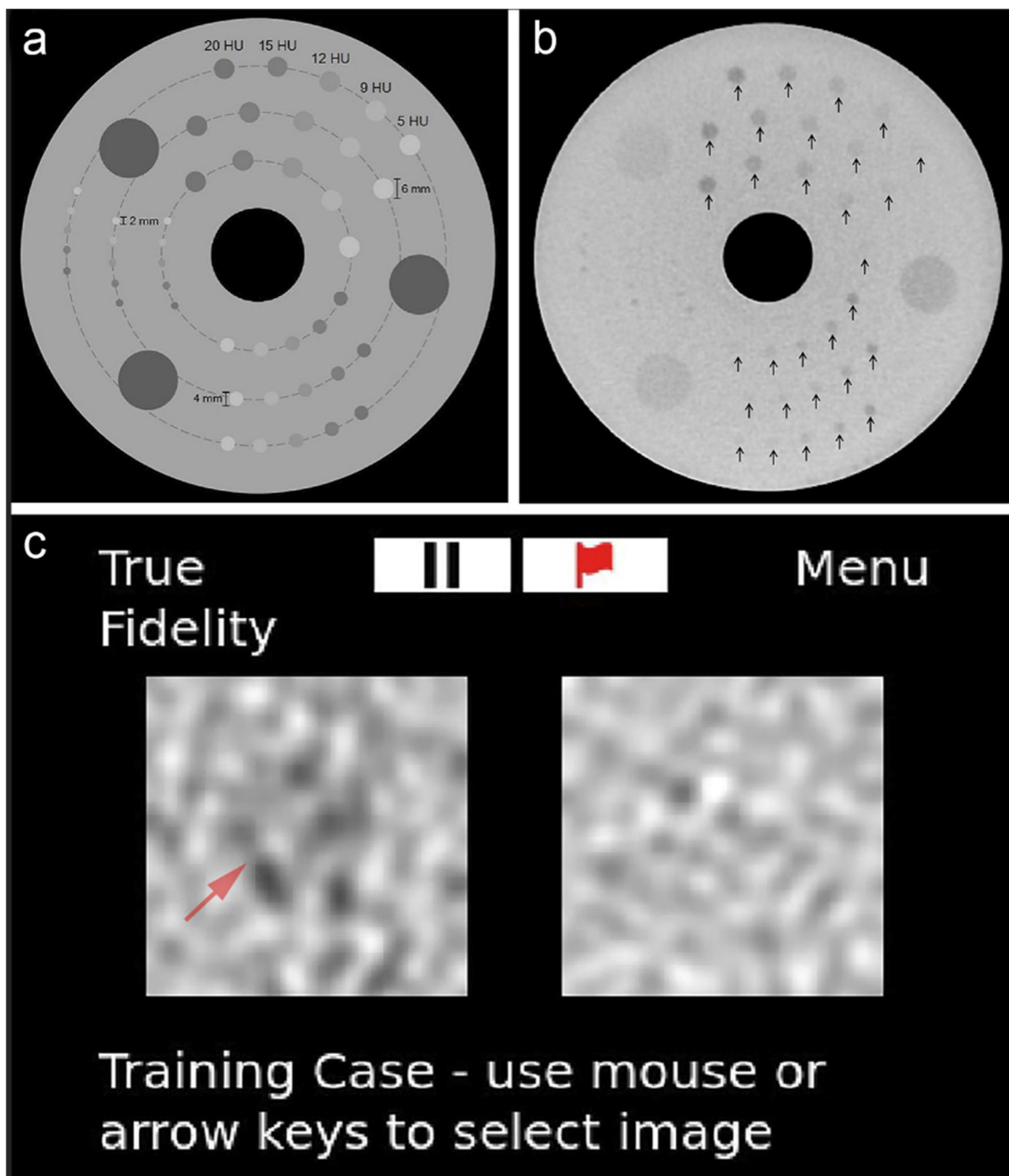
Fig. 1 Schematic illustration (**a**) and CT image (**b**) of the contrast-detail phantom. Forty-five low-contrast lesions consisting of five contrast levels at 120 kVp and three sizes are arranged in three groups. A screenshot of our two-alternative forced choice experiment with a lesion (pink arrow) in the left image and no lesion in the right image (**c**)

experience in image quality assessment, performed the image interpretation blindly. Each reader read a total of 1050 2AFC trials, each consisting of two side-by-side images: lesion-present (i.e., the object) and lesion-absent (i.e., background noise) (Fig. 1c). The lesion-present and lesion-absent images represented the circular regions of interest taken from the cylindrical objects and uniform regions in the phantom, respectively. These images were randomly selected from all possible combinations of all dose indexes, investigated reconstruction algorithms, and lesion size (4 mm and 6 mm). The contrast level used was 20 HU because of its appropriate level of subtlety. Image pairs were shown in randomized order. Each reader was instructed to select the image of the pair that was most likely to contain a lesion. The results (detected or not detected) were recorded and the 2AFC analysis was carried out following previously described methods [16, 18].

Table 1 Scan parameters for phantom experiment and clinical study

Parameter	Phantom	Clinical
Scanner mode	Revolution (GE Healthcare)	Revolution (GE Healthcare)
Scan mode	Axial	Helical
Beam Collimation (mm)	100	80
Tube voltage (kVp)	120	120
Tube current (mA)	335, 200, 100, 70, 35	Full dose (400) / Low dose (280/200/120)
Rotation (s)	0.5	0.5
Pitch	/	0.992:1
SFOV (cm)	Small body	Large body
DFOV (cm)	22	50
CT dose index (mGy)	10, 6, 3, 2, 1	13.6, 9.5, 6.8, 4.1
Section thickness/interval (mm)	2.5/2.5	2.5/2.5
Reconstruction algorithm	FBP, IR(ASiR-V60%), DLIR at medium level	FBP, IR(ASiR-V60%), DLIR at medium level
Reconstruction kernel	Standard	Standard

Note. *DFOV*, display field of view; *SFOV*, scan field of view; *FBP*, filtered back projection; *IR*, iterative reconstruction; *DLIR*, deep learning image reconstruction; *ASiR-V*, hybrid model-based adaptive statistical iterative reconstruction

Clinical study

Participants and study design

The institutional review board approved this single-institution prospective study, and all participants provided written informed consent before enrollment. One team member was partially funded through a GE Research Fellowship Scholarship and one author of the study (L.W.) is an employee of GE Healthcare China. Those authors (non-consultants for the company) controlled the study data.

This investigator-initiated study was a three-group trial that assessed two non-inferiority hypotheses in allcomers: (1) that DLIR at reduced radiation level (test) was non-inferior to FBP at full dose (100% dose, reference), and (2) that DLIR at reduced radiation levels was non-inferior to IR at full dose (100% dose, control) when it comes to the primary task (image quality and detection of hepatic metastases). Reduced dose levels of DLIR (70%, 50%, and 30% of the full dose) were obtained from three groups of adult participants (Fig. 2). These dose settings were determined based on the above phantom study and volume CTDI (CTDI$_{vol}$) levels from abdominopelvic CT scans reported by the College of Radiology Dose Index Registry [19].

From October 2020 to April 2021, 190 consecutive participants who underwent clinically indicated, contrast material–enhanced abdominopelvic CT scan for hepatic metastases were prospectively enrolled. The inclusion-exclusion criteria are outlined in Fig. 2. The final study sample included 154 participants, who were divided into three groups based on the study period and low-dose scanning method [20, 21]. There were 54 participants in group A (70% dose), 50 participants in group B (50% dose), and 50 participants in group C (30% dose) during the first, middle, and last third of the study period, respectively.

Image acquisition and radiation dose

All CT scans were done on the Revolution CT scanner. Nonionic contrast material (Iohexol, Omnipaque, 350 mg I/mL; GE Healthcare) was administered intravenously at a rate of 3 mL/s for a total of 70–120 mL (1.2 mL/kg). Following the acquisition of the late arterial phase, which began 15 s after aortic peak attenuation on the timing bolus, two portal venous phase passes were performed at about 60 s (initial diagnostic scan) and 66 s (low-dose portal venous scans), respectively (Table 1) [22]. The arterial phase was not evaluated in this investigation. There were a total of six data sets per participant (one group, two radiation dose levels, and three reconstruction algorithms), which led to a total of 924 data sets for the entire study population.

Each participant's radiation dose data, including CTDIvol and dose-length product, was recorded. Based on the method described in American Association of Medical Physicists Report 204 [23], a size-specific dose estimate was calculated by multiplying the CTDI$_{vol}$ by a size-dependent conversion factor.

Image analysis

Qualitative analysis Under standard clinical conditions, subjective image quality was independently assessed by three
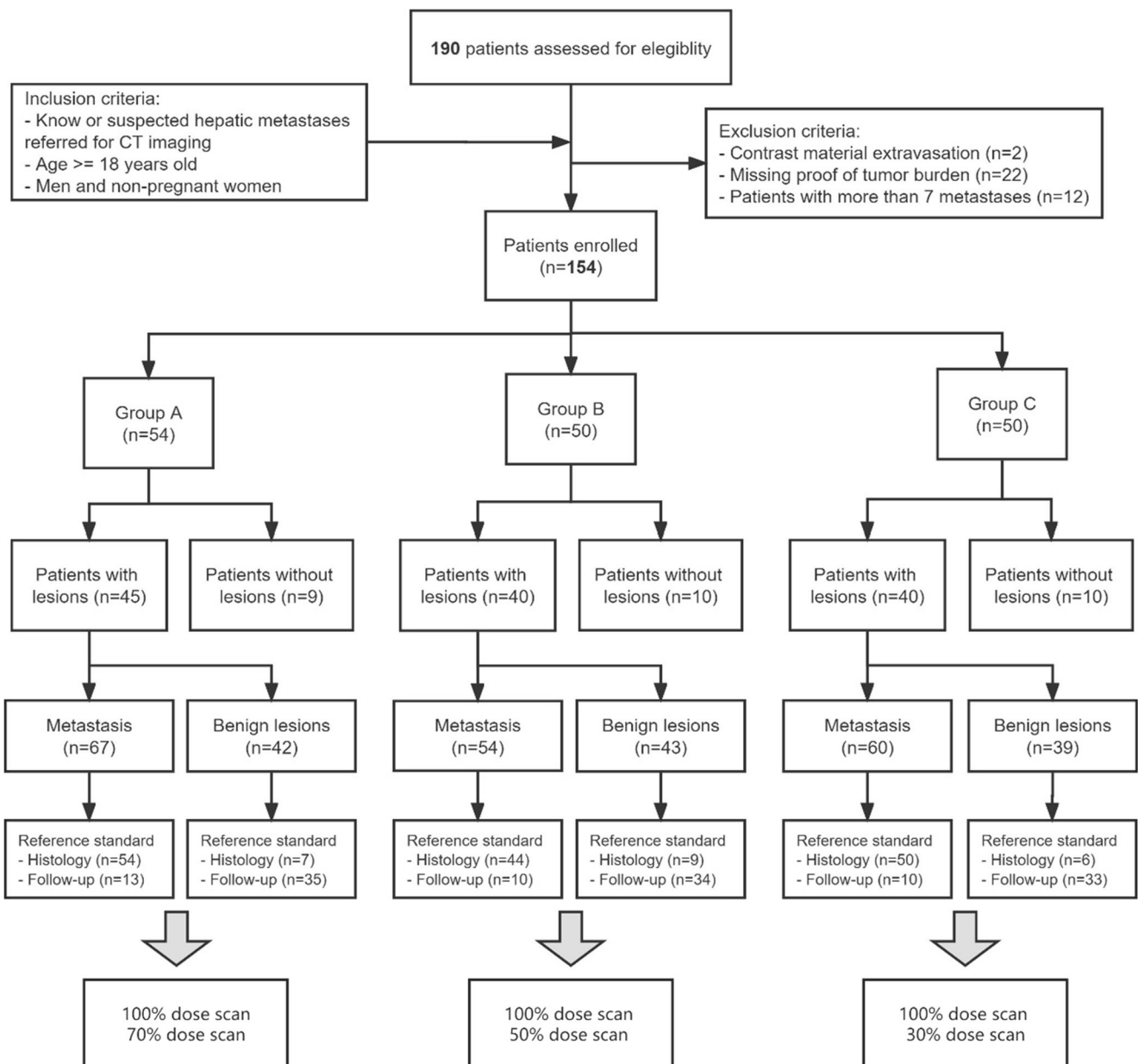
Fig. 2 Flowchart of study population enrollment

independent radiologists with different years of experience in abdominal imaging (3–11 years of experience) using a dedicated radiology imaging viewer software (RadiAnt DICOM Viewer v.5.0.2). All image datasets were anonymized to remove patient identifiable information before being presented to the readers. The readers were allowed to scroll, pan/zoom, and adjust window settings during reading of subjective image quality evaluation and lesion detection task without any time limit. Using a five-point ordinal scale, readers were first instructed to rank the noise, distortion, sharpness, and overall quality of images according to their expectations for the detection of hepatic lesions (1 = very poor, 2 = poor, 3 = average, 4 = above average, and 5 = excellent) using the methods described

in previous reports [24]. Readers were told that an image scoring less than or equal to 2 would be deemed inadequate for diagnosis (Appendix E1).

Lesion analysis and reference standard Following the image quality assessment, the same readers were asked to determine the presence or absence of hepatic lesions, and then to rate the conspicuity of the identified lesions based on a five-point scale (1 = unreadable, 2 = poor, 3 = moderately certain, 4 = good, and 5 = excellent). The readers were then asked to differentiate between metastases and benign lesions, and to assign a lesion-level confidence score, which indicated the confidence related

to the primary task (the detection of hepatic metastases), using a five-point confidence index (1, very poor; 2, poor; 3, average; 4, above average; and 5, excellent confidence). If a lesion was considered benign which was not related to the primary task, it would be assigned a low numeric confidence score from 1 to 2.

Reference standards for lesions were established on routine-dose IR images by two nonblinded consensus radiologists (35 years and 14 years of experience, respectively). Lesions were identified and classified as malignant or benign based on a combination of imaging (e.g., MRI, PET-CT or follow-up) and pathology (biopsy and surgery). The follow-up interval at which benign lesions were considered was over 6 months. Participants with no hepatic lesions were identified based on the clinical data and subsequent or previous CT or MRI imaging for at least 8 months (range, 8–12 months). In the case that any reader detected a lesion at a specific dose, that lesion was considered detected. As for the classification of lesions, it should only be considered accurate when two or more readers classify the lesion as either malignant or benign against the reference standard.

Statistical analysis

The results from the 2AFC detection experiment were analyzed using a generalized linear mixed-effects model, with the goodness of fit determined by using AIC, BIC, and log likelihood. Pairwise comparisons between two reconstruction algorithms were performed with the 95% confidence interval (CI) calculated. The method of determining the dose reduction potential of the DLIR algorithm, follows a previously published technique [16].

We estimated that 50 participants in each group would provide 80% power or higher to demonstrate non-inferiority with a margin of image quality (−0.5 score) and a 2.5% one-sided significance level [25] (PASS 16, NCSS Statistical Software). Analyses were performed on the basis of intention-to-treat. A jackknife alternative free-response receiver operating characteristic (JAFROC) figure of merit (FOM) non-inferiority analysis was performed to compare the reader performance for detecting hepatic metastasis between routine-dose CT and lower-dose CT configurations. Comparisons of FOMs were estimated by using the RJafroc package v1.0.1 (R version 3.4.2) for R. A positive participant case containing multiple hepatic metastases was weighted based on the reciprocal. Prior to the study, the noninferiority margin for the calculation of the difference between the routine and lower dose configurations was set at −0.10, such that the lower limit of the 95% CI had to be greater than −0.10 [6, 25].

Generalized estimating equations (GEE) with independent working covariance matrices were utilized in the qualitative image analysis. Inter-reader agreement was assessed using Fleiss' kappa statistics. In order to determine whether lesion

size (diameter) or liver-to-lesion contrast-to-noise ratio (CNR) affected metastasis detection, correlations to mean reader detection confidence at each dose and reconstruction were examined by Spearman rank correlation coefficient [6].The liver-to-lesion CNR was obtained by dividing the CT difference between a metastasis and the background liver by the background liver CT noise (Appendix E1).

Results

Phantom experiment

A subset of reconstructed phantom images is shown in Fig. 3. The detection accuracy ranged from 69% (95% CI: 61%, 77%) to 97% (95% CI: 95%, 99%) and generally increased with increasing dose index and object size. The average of 95% CI was 10% (median, 10%; interquartile range [IQR], 8–13%) representing interobserver variability. DLIR and IR did not differ statistically, but both were better than FBP in detection accuracy (both $p < 0.001$) with a mean absolute difference of 3.0 (95% CI: 1.4%, 4.6%) for the former and 3.1 (95% CI: 1.3%, 4.8%) for the latter. The dose reduction potential from DLIR was estimated to be between 13% (95% CI: 8%, 15%) and 57% (95% CI: 34%, 61%) (average ± 95% CI, 31% ± 25) based on the reference FBP dose index for both 4 mm and 6 mm object sizes (Fig. 4).

Clinical study

The demographic characteristics of the three groups did not differ significantly (Table 2). A total of 54 participants had 109 lesions in group A (100% and 70% doses), 50 participants had 97 lesions in group B (100% and 50% doses), and 50 participants had 99 lesions in group C (100% and 30% doses). The mean size of 181 metastases and 124 benign lesions was 1.5 cm ± 0.8 and 1.6 cm ± 1.1, respectively.

Readers' scores for all image quality parameters and all reconstruction methods gradually decreased as radiation dose was gradually reduced (Fig. 5). DLIR consistently outperformed FBP and IR for all radiation dose levels ($p < 0.01$ for all comparisons), with substantial and perfect agreement (kappa values of 0.73–0.85 [95% CI of 0.62–0.96]). Furthermore, the overall percentage of FBP and IR examinations that did not meet reasonable quality standards was 5.6% (52/924) and 2.5% (23/924), respectively, but only 0.9% (8/924) for DLIR ($p < 0.001$) (Table E1). In general, DLIR's overall image quality was comparable to that of full-dose FBP and IR except that the image quality of DLIR with 70% radiation dose reduction was inferior to full-dose IR ($p < 0.001$) (Table 3 and Fig. E1).

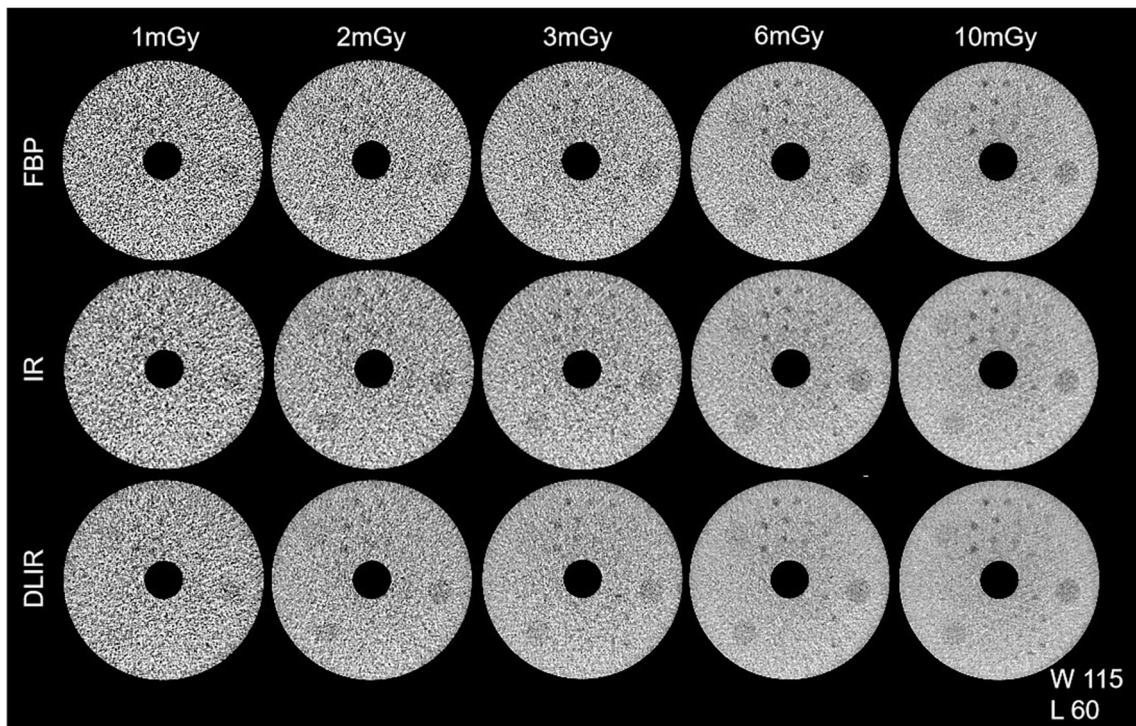At the same dose level, no significant differences in detecting hepatic lesions were observed between the three

Fig. 3 Reconstructed CT images of the contrast-detail phantom obtained with three reconstruction algorithms and five dose levels at 2.5-mm section thickness. FBP, filtered back projection; IR, iterative reconstruction; DLIR, deep learning imaging reconstruction

reconstructions (Table E2). A comparison of low-dose DLIR and full-dose FBP/IR in detecting hepatic lesions revealed no significant difference, with the exception that 30% dose DLIR had a lower accuracy in detecting small lesions (< 1 cm) than full-dose FBP/IR ($p < 0.001$). Table 4 provides the estimated difference in JAFROC FOMs of DLIR at reduced radiation levels from FBP at full dose and IR at full dose, respectively. Non-inferior performance for detecting hepatic metastases
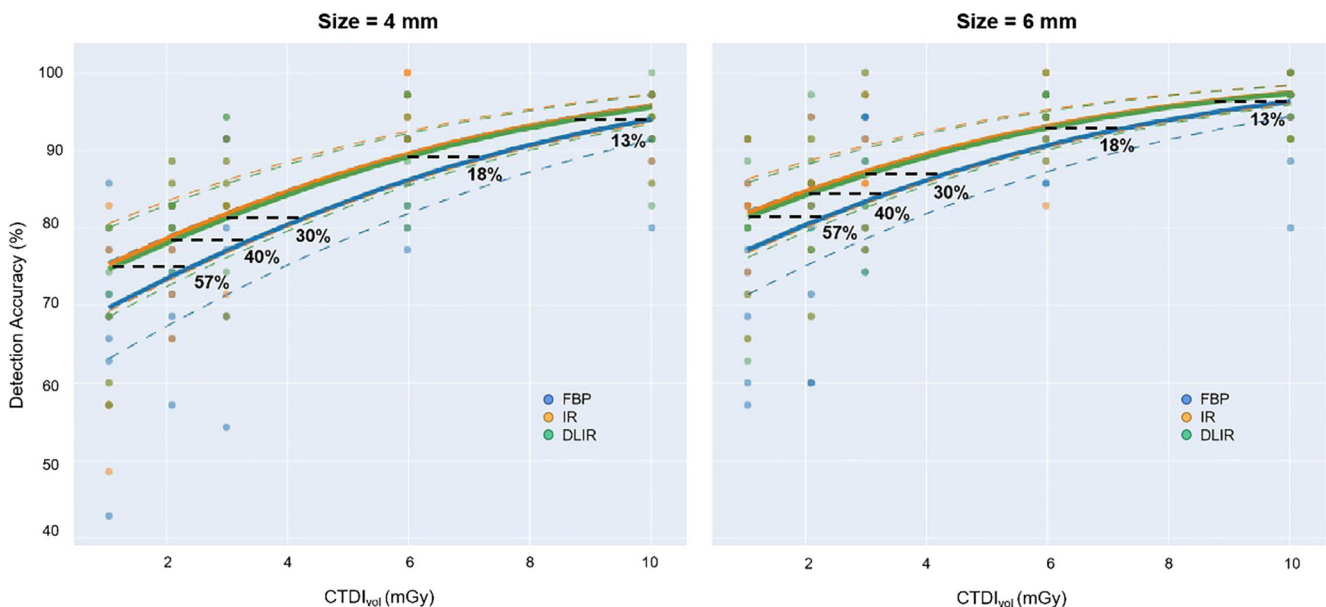

Fig. 4 Plots showing average detection accuracy as a function of radiation dose for filtered back projection (FBP), iterative reconstruction (IR), and deep learning imaging reconstruction (DLIR). Solid lines represent detection accuracy averaged across readers, and dashed lines represent radiation dose reduction potential of DLIR over the different reference FBP dose index levels. Small dots represent individual readers

Table 2 Patient demographics and breakdown of hepatic metastasis

Parameter	Total ($n = 154$)	Group A ($n = 54$)	Group B ($n =50$)	Group C ($n =50$)	p value[#]
No. of men (%)*	92 (59.7)	32 (59.2)	32 (64.0)	28 (56.0)	0.21
Age (y)	59 ± 12	61 ± 10	59 ± 13	58 ± 13	0.61
Men (y)	60 ± 12	61 ± 11	60 ± 13	59 ± 14	0.85
Women (y)	58 ± 11	61 ± 10	57 ± 14	56 ± 11	0.77
Weight (kg)	64.9 ± 13.3	63.7 ± 12.7	66.2 ± 14.5	63.9 ± 13.0	0.43
Body mass index	23.5 ± 4.0	23.2 ± 3.5	23.7 ± 3.8	23.8 ± 4.8	0.79
Anterior-posterior diameter (cm)	22.6 ± 3.6	23.6 ± 4.0	23.2 ± 3.7	21.4 ± 3.2	0.20
Transverse diameter (cm)	29.7 ± 3.0	30.3 ± 3.9	30.5 ± 3.3	28.7 ± 3.6	0.15
Effective diameter (cm)	26.1 ± 2.9	26.8 ± 3.5	26.9 ± 3.1	24.9 ± 3.0	0.13
Full-dose DLP (mGy.cm)	547.1 ± 161.05	552.3 ± 189.0	547.1 ± 135.2	541.2 ± 139.8	0.76
Full-dose SSDE (mGy)	19.3 ± 2.0	19.0 ± 2.4	18.6 ± 2.4	20.2 ± 2.3	0.11
Low-dose DLP (mGy.cm)	278.7 ± 101.1	386.6 ± 122.26	273.5 ± 87.62	162.4 ± 61.93	**< 0.0001**
Low-dose SSDE (mGy)	9.2 ± 3.2	13.3 ± 3.8	9.3 ± 2.8	6.1 ± 1.7	**< 0.0001**
Type of metastasis (No. of lesions/No. of patients)	181/66	67/24	54/20	60/22	0.74/0.99
Colorectal	66/22	19/6	19/7	28/9	
Lung	33/13	13/5	11/4	9/4	
Stomach	23/9	10/4	8/3	5/2	
Esophagus	15/5	7/2	3/1	5/2	
Pancreas	8/4	4/2	2/1	2/1	
Other	36/13	14/5	11/4	11/4	
Type of benign lesions (No. of lesions/No. of patients)	124/59	42/21	43/20	39/18	0.61/0.98
Hemangioma	12/7	5/3	4/2	3/2	
Cysts	91/42	33/15	30/14	28/13	
Other	21/10	4/3	9/4	8/3	
Lesion size (cm)	1.6 ± 1.0	1.6 ± 0.9	1.7 ± 1.1	1.4 ± 0.9	0.22
Metastasis	1.5 ± 0.8	1.3 ± 0.9	1.7 ± 1.0	1.3 ± 0.8	0.62
Benign lesions	1.6 ± 1.1	1.7 ± 1.0	1.6 ± 1.2	1.5 ± 0.9	0.23

Note. Data are mean ± standard deviation unless otherwise indicated. *Data are numbers with percentages in parentheses. [#] Significant differences between groups at $p < 0.05$. *BMI*, body mass index; *CTDIvol*, volume CT dose index; *DLP*, dose-length product; *SSDE*, size-specific dose estimation

was obtained at 50% and 70% doses for DLIR. However, inferior performance was found at 30% dose for DLIR in detection of small lesions (< 1 cm) than the 100%-dose FBP (difference: −0.112; 95% CI: −0.178 to 0.047) ($p < 0.001$) and 100%-dose IR (difference: −0.123; 95% CI: −0.182 to 0.053) ($p < 0.001$) (Fig. 6, Fig. E2-E4). For each radiation dose level, the sensitivity and specificity of GEE per lesion for metastasis were not significantly different between the three reconstructions (Tables E3 and E4); however, the mixed-effects logistic regression model showed DLIR to be more sensitive than FBP for the combined data at different radiation doses (odds ratio = 1.45, 95% CI: 1.03, 2.05, $p = 0.035$).

In addition, liver-to-lesion CNR and readers' perception of lesion conspicuity, as well as reader confidence, were improved with DLIR over FBP and IR across all investigated radiation dose levels ($p < 0.05$). Lesion size and liver-to-lesion CNR were correlated with reader confidence ($p < 0.05$). No significant differences were found in the correlations for these parameters between these three reconstructions.

Discussion

The results from our phantom experiment and three-arm non-inferiority clinical study are largely consistent and support the notion that DLIR could achieve similar performance to FBP or IR at reduced dose levels for low-contrast detection tasks. Low-dose DLIR performed non-inferiorly to full-dose FBP at dose reductions of 70% and to full-dose IR at dose reductions of 50% in terms of image quality and the detection of hepatic metastases (≧ 1 cm). In contrast, DLIR was non-inferior for small lesions (< 1 cm) at dose reductions of no greater than 50% compared to full-dose FBP/IR while maintaining non-inferior image quality.

Previous studies have shown DLIR improves image quality when compared to FBP or IR, but its ability to facilitate dose reduction for low-contrast tasks is less clear [5, 17, 26]. Low-contrast detection tasks are mainly affected by image contrast differences. Our study demonstrated that DLIR consistently improved image noise, image quality, and metastasis CNR in
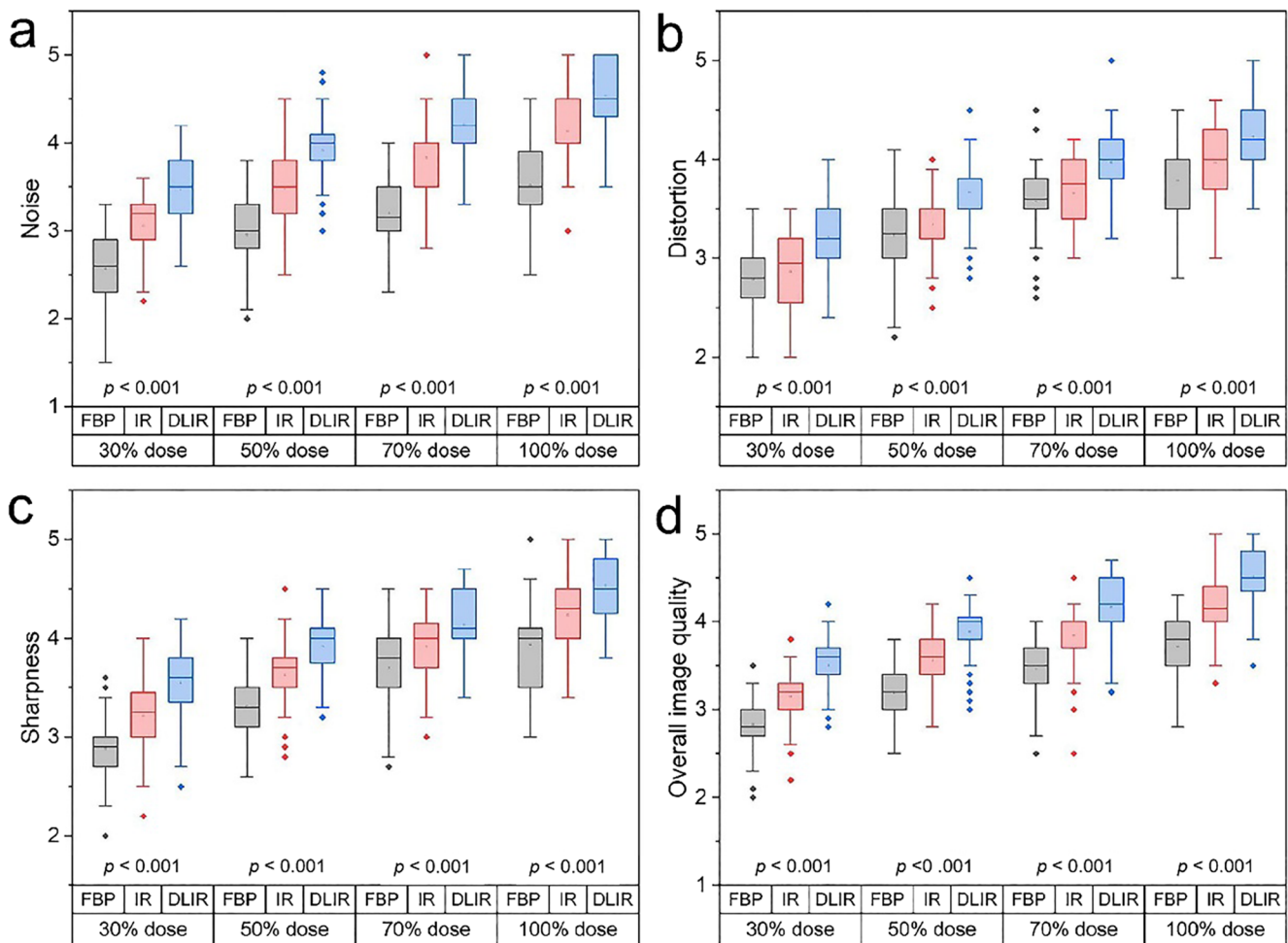
Fig. 5 Box plots of (**a**) noise, (**b**) distortion, (**c**) sharpness, and (**d**) overall image quality as a function of dose level and reconstruction algorithm. The generalized estimating equations were used for comparing image quality metrics between reconstruction algorithms at every dose level (*p* values at the bottom of the plots). At each dose level, the three reconstruction algorithms differ significantly for each parameter, with comparison to FBP and IR across a range of dose levels. With these improvements, DLIR was able to detect hepatic metastases at reduced dose levels with increased overall sensitivity and higher reader confidence. However, DLIR's radiation-saving potential is limited due to readers' declining scores on all image deep learning imaging reconstruction outperforming filtered backprojection and iterative reconstruction. Also note that higher values on the y-axis indicate lower perceived noise levels (**a**). FBP, filtered back projection; IR, iterative reconstruction; DLIR, deep learning imaging reconstruction

quality parameters and lesion assessment with decreasing dose levels, as with standard reconstructions. Furthermore, lesions with less CNR and smaller sizes were associated with reduced reader confidence. These results might explain why the dose reduction of DLIR in detecting hepatic metastasis (irrespective

Table 3 Noninferiority test results of low-dose DLIR compared with full-dose FBP and IR for overall image quality

Group	Full-dose FBP	Full-dose IR	Low-dose DLIR[#]	DLIR vs. FBP*	DLIR vs. IR*
A	3.72 ± 0.37 (3.63, 3.81)	4.15 ± 0.34 (4.07, 4.23)	4.16 ± 0.32 (4.09, 4.23)	0.44 (0.31, 0.56)	0.01 (−0.12, 0.13)
B	3.75 ± 0.46 (3.65, 3.86)	4.21 ± 0.38 (4.12, 4.33)	3.89 ± 0.30 (3.82, 3.96)	0.14 (0.05, 0.29)	−0.32 (−0.42, −0.14)
C	3.68 ± 0.28 (3.58, 3.72)	4.12 ± 0.31 (3.98, 4.20)	3.50 ± 0.29 (3.44, 3.57)	−0.18 (−0.34, −0.10)	**−0.62 (−0.77, −0.53)**

Note. Data are expressed as mean ± SD, with 95% confidence intervals in parentheses. *Mean difference between low-dose DLIR compared to full-dose FBP or full-dose IR. [#] Low doses for DLIR in groups A, B, and C were 70% dose, 50% dose, and 30% dose respectively. The three readers showed substantial and perfect agreement on lesion characterization at each dose level and size (kappa values, 0.73–0.85 [95% CI: 0.62–0.96]). *FBP*, filtered back projection; *IR*, iterative reconstruction; *DLIR*, deep learning imaging reconstruction

Table 4 Pooled JAFROC FOMs and comparisons for hepatic metastases

Size	Group	Average JAFROC FOM			DLIR vs. FBP*	DLIR vs. IR*
		Full-dose FBP	Full-dose IR	Low-dose DLIR[#]		
< 1 cm	A (n = 71)	0.901 (0.823, 0.979)	0.908 (0.827, 0.979)	0.886 (0.790, 0.949)	−0.015 (−0.010, 0.016)	−0.022 (−0.026, 0.027)
	B (n = 61)	0.884 (0.819, 0.973)	0.898 (0.794, 0.961)	0.844 (0.753, 0.939)	−0.040 (−0.075, 0.033)	−0.054 (−0.082, 0.013)
	C (n = 56)	0.881 (0.767, 0.952)	0.892 (0.780, 0.958)	0.769 (0.637, 0.871)	**−0.112 (−0.178, 0.047)**	**−0.123 (−0.182, 0.053)**
≧ cm	A (n = 38)	0.962 (0.853, 0.997)	0.977 (0.868, 1.000)	0.948 (0.821, 0.995)	−0.014 (−0.056, 0.005)	−0.029 (−0.030, 0.008)
	B (n = 36)	0.968 (0.855, 0.999)	0.972 (0.855, 0.999)	0.944 (0.813, 0.993)	−0.024 (−0.027,0.009)	−0.028 (−0.032,0.010)
	C (n = 44)	0.955 (0.843, 0.995)	0.963 (0.854, 0.998)	0.930 (0.809, 0.985)	−0.024 (−0.038,0.012)	−0.033 (−0.053,0.016)

Note. The pooled jackknife alternative free-response receiver operating characteristic (JAFROC) figures of merit (FOMs) represented in this table were obtained from across three abdominal radiologists for the detection of hepatic metastases. Data in parentheses are 95% confidence intervals. *Difference between FOMs in low-dose DLIR compared to full-dose FBP or full-dose IR. [#] Low doses for DLIR in groups A, B, and C were 70% dose, 50% dose, and 30% dose, respectively. *FBP*, filtered back projection; *IR*, iterative reconstruction; *DLIR*, deep learning imaging reconstruction
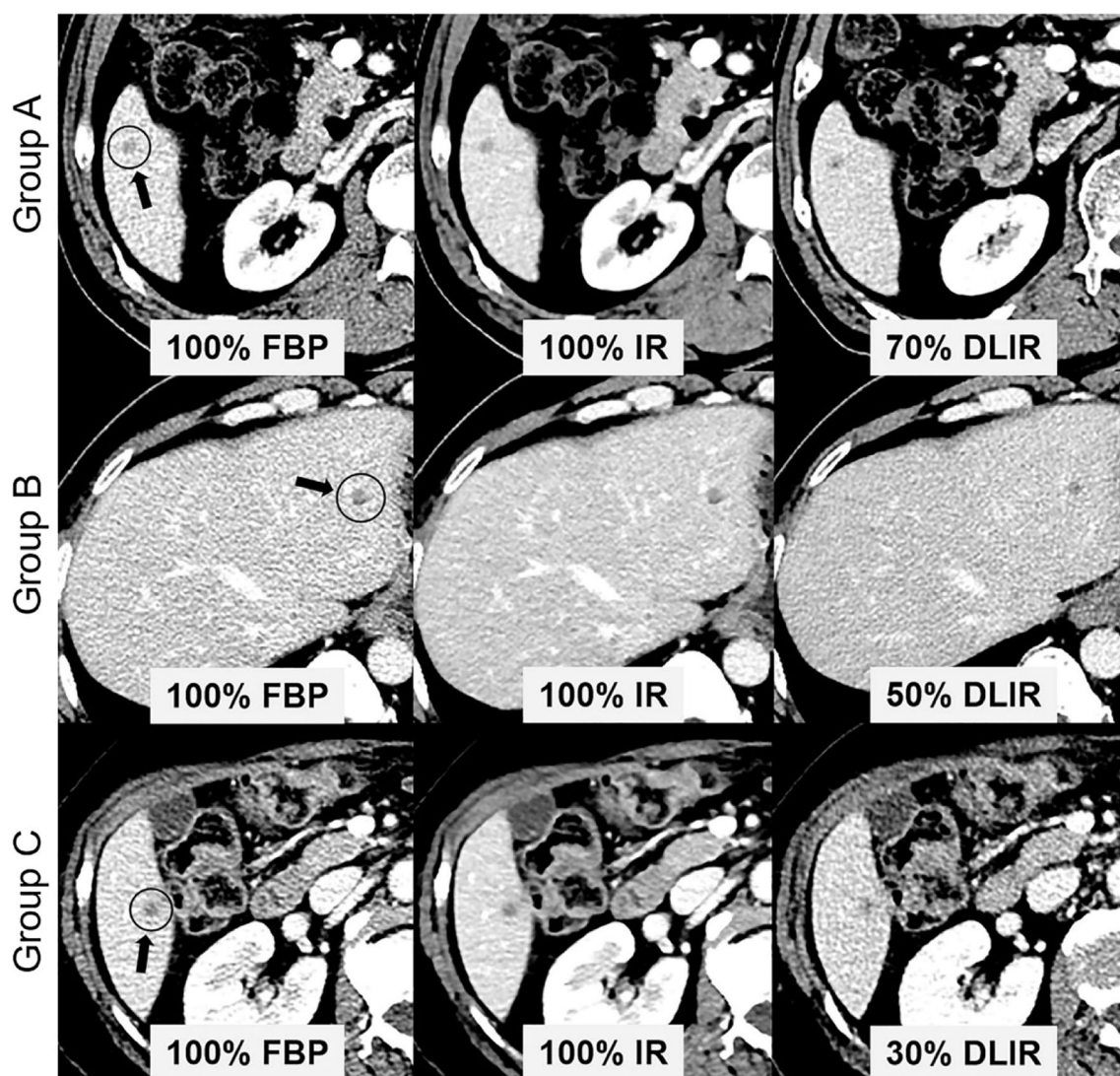

Fig. 6 Axial contrast-enhanced CT images of the abdomen obtained with 100%-dose filtered back projection (FBP), 100%-dose iterative reconstruction (IR), and low-dose deep learning image reconstruction (DLIR) in the same breath hold. All liver metastases (arrows with circles; 0.5 cm, 0.7 cm, and 0.5 cm in groups A, B, and C) were detected by all readers at 100%-dose FBP, 100%-dose IR, 70%- and 50%-dose DLIR. The liver metastases at 30%-dose DLIR had a contrast-to-noise ratio of 2.1, which was missed by one reader

of tumor size) could not exceed 50% of the full-dose FBP/IR. A larger practical benefit of DLIR is that it is likely to help radiologists accept lower-dose abdominal CT images due to its perceived image quality improvement over FBP and IR.

The extent to which DLIR can reduce radiation doses will depend on the references, reconstruction settings, and primary clinical tasks intended. Our 2AFC detection experiment indicated that medium-strength DLIR reduced dose by up to 57% when compared with FBP, slightly less than the 67% reduction obtained by Racine et al [27] when comparing the detectability index of high-strength DLIR and FBP. Clinically, DLIR was inferior to FBP/IR at a 70% dose reduction for detecting small hepatic metastases (< 1 cm), which aligns well with Jensen et al [14] who reported inferior detection of liver metastases (< 0.5 cm) using DLIR at 65% dose reduction. The difference in our study is that we also examined the effectiveness of DLIR dose reduction in maintaining the perception of image quality. The overall image quality of DLIR was largely diagnostically acceptable despite the 70% dose reduction (95%, 142/150), which was comparable to full-dose FBP, although it was inferior to full-dose IR. These findings are similar to the previous study [15] which showed radiation dose reduction of > 75% of DLIR in the whole-body CT while maintaining comparable image quality and detection rate of systemic lesions in comparison to standard dose-IR (CTDI$_{vol}$: 2.9 mGy vs. 13.5 mGy).

While this study focused exclusively on the DLIR algorithm of one specific scanner model (GE Apex), its practical implications were considerable. Our multireader- and multidose-based research results provide strong evidence for the feasibility of low-dose DLIR application in routine clinical practice, which will help to reduce the potential radiation exposure damage caused by multiple abdominal CT re-examinations of oncological patients. In addition, we did not find any research papers in the literature that examined the correlations between the new DLIR algorithm and standard algorithms across multiple doses in humans. Furthermore, our low-dose DLIR results in abdominal CT may serve as a valuable reference for the low-dose applications of other commercially available deep learning–based CT denoising algorithms, such as Canon's Advanced Intelligent Clear-IQ Engine [12, 28] and Neusoft's NeuAI denosing [29]. Dose reduction reported by these algorithms varied greatly from approximately 30 to 80% with regard to maintaining image quality and diagnostic value (e.g., hepatic lesion detection). When implementing new technology into routine oncologic CT imaging, it is necessary to strike a balance between radiation dose, image quality, and clinical task. Considering that liver evaluation of potential metastatic disease has a higher priority than radiation protection, we cautiously suggest a 50% dose reduction (CTDIvol of 6.8 mGy) of DLIR rather than a 70% dose reduction (4.1 mGy) in routine oncologic imaging.

There were several limitations to our study. First, as a relatively large population and an investigation conducted in one institution, individual practice habits and preferences might affect the results. To reduce the effects of interreader variability, we performed multireader image evaluation for phantom (eleven readers) and clinical trials (three readers). It was found that the interreader variabilities for both phantom and clinical trials were low, which made our results to be credible and generalizable. Second, the comparisons between the full-dose and low-dose CT scans were separated into three groups due to ethical considerations of radiation dose. We believed the results were not affected, however, since the comparisons were mainly conducted within the same participants, and the participants' demographics and baseline characteristics were homogeneous across groups. Third, studies with human observers were inevitably constrained by human factors such as reader fatigue; consequently, only a small number of key reconstructions were evaluated. Fourth, all participants were of a single ethnicity, meaning ethnicity-based image analyses cannot be assessed.

In conclusion, the liver-to-lesion CNR, readers' perception of image quality, and lesion conspicuity, as well as reader confidence, were significantly improved with DLIR over FBP and IR across all investigated radiation dose levels. DLIR allowed for a 50% dose reduction for detecting low-contrast hepatic metastases (irrespective of tumor size) while maintaining comparable image quality to full-dose FBP and full-dose IR.

Declarations

Guarantor The scientific guarantor of this publication is Dr. Jianbo Gao.

Conflict of interest The authors of this manuscript declare relationships with the following companies: GE Healthcare China for Luotong Wang. The other authors declare that they have no conflict of interest.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent The institutional review board approved this single-institution prospective study (Registry number: ChiCTR-DPD-16010302), and all participants provided written informed consent before enrollment.

Ethical approval   Institutional Review Board approval was obtained.

Methodology
- prospective
- diagnostic study
- performed at two institutions (phantom part performed in Duke University Medical Center, clinical part performed in The First Affiliated Hospital of Zhengzhou University)

References

1. Geyer LL, Schoepf UJ, Meinel FG et al (2015) State of the art: iterative CT reconstruction techniques. Radiology 276:339–357
2. Mileto A, Guimaraes LS, McCollough CH, Fletcher JG, Yu L (2019) State of the art in abdominal CT: the limits of iterative reconstruction algorithms. Radiology 293:491–503
3. Solomon J, Lyu P, Marin D, Samei E (2020) Noise and spatial resolution properties of a commercially available deep learning-based CT reconstruction algorithm. Med Phys 47:3961–3971
4. Chen B, Christianson O, Wilson JM, Samei E (2014) Assessment of volumetric noise and resolution performance for linear and non-linear CT reconstruction methods. Med Phys 41:071909
5. Nagayama Y, Sakabe D, Goto M et al (2021) Deep learning-based reconstruction for lower-dose pediatric CT: technical principles, image characteristics, and clinical implementations. Radiographics 41:1936–1953
6. Fletcher JG, Fidler JL, Venkatesh SK et al (2018) Observer performance with varying radiation dose and reconstruction methods for detection of hepatic metastases. Radiology 289:455–464
7. McCollough CH, Yu L, Kofler JM et al (2015) Degradation of CT low-contrast spatial resolution due to the use of iterative reconstruction and reduced dose levels. Radiology 276:499–506
8. Choi H, Chang W, Kim JH et al (2022) Dose reduction potential of vendor-agnostic deep learning model in comparison with deep learning-based image reconstruction algorithm on CT: a phantom study. Eur Radiol 32:1247–1255
9. Racine D, Brat HG, Dufour B et al (2021) Image texture, low contrast liver lesion detectability and impact on dose: deep learning algorithm compared to partial model-based iterative reconstruction. Eur J Radiol 141:109808
10. Greffier J, Hamard A, Pereira F et al (2020) Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. Eur Radiol 30:3951–3959
11. Hsieh J, Liu E, Nett B, Tang J, Thibault J-B, Sahney S (2019) A new era of image reconstruction: TrueFidelity™. White Paper (JB68676XX), GE Healthcare
12. Bornet PA, Villani N, Gillet R et al (2022) Clinical acceptance of deep learning reconstruction for abdominal CT imaging: objective and subjective image quality and low-contrast detectability assessment. Eur Radiol 32:3161–3172
13. Fletcher JG, Yu L, Fidler JL et al (2017) Estimation of observer performance for reduced radiation dose levels in CT: eliminating reduced dose levels that are too low is the first step. Acad Radiol 24:876–890
14. Jensen CT, Gupta S, Saleh MM et al (2022) Reduced-dose deep learning reconstruction for abdominal CT of liver metastases. Radiology 303:90–98
15. Noda Y, Kaga T, Kawai N et al (2021) Low-dose whole-body CT using deep learning image reconstruction: image quality and lesion detection. Br J Radiol :20201329
16. Solomon J, Mileto A, Ramirez-Giraldo JC, Samei E (2015) Diagnostic performance of an advanced modeled iterative reconstruction algorithm for low-contrast detectability with a third-generation dual-source multidetector CT scanner: potential for radiation dose reduction in a multireader study. Radiology 275:735–745
17. Jensen CT, Liu X, Tamm EP et al (2020) Image quality assessment of abdominal CT by use of new deep learning image reconstruction: initial experience. AJR Am J Roentgenol 215:50–57
18. Solomon J, Marin D, Roy Choudhury K, Patel B, Samei E (2017) Effect of radiation dose reduction and reconstruction algorithm on image noise, contrast, resolution, and detectability of subtle hypoattenuating liver lesions at multidetector CT: filtered back projection versus a commercial model-based iterative reconstruction algorithm. Radiology 284:777–787
19. Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL (2017) U.S. diagnostic reference levels and achievable doses for 10 adult CT examinations. Radiology 284:120–133
20. Lv P, Zhou Z, Liu J et al (2019) Can virtual monochromatic images from dual-energy CT replace low-kVp images for abdominal contrast-enhanced CT in small- and medium-sized patients. Eur Radiol 29:2878–2889
21. Lee KH, Lee JM, Moon SK et al (2012) Attenuation-based automatic tube voltage selection and tube current modulation for dose reduction at contrast-enhanced liver CT. Radiology 265:437–447
22. Bellini D, Ramirez-Giraldo JC, Bibbey A et al (2017) Dual-source single-energy multidetector CT used to obtain multiple radiation exposure levels within the same patient: phantom development and clinical validation. Radiology 283:526–537
23. Boone JMSK, Cody DD, McCollough CH, McNitt-Gray MF, Toth TL (2011) Size-specific dose estimates (SSDE) in pediatric and adult body CT examinations. In: Report of Am Assoc Phys Med AAPM Task Group 204. American Association of Physicists in Medicine, College Park
24. Nam JG, Hong JH, Kim DS, Oh J, Goo JM (2021) Deep learning reconstruction for contrast-enhanced CT of the upper abdomen: similar image quality with lower radiation dose in direct comparison with iterative reconstruction. Eur Radiol 31:5533–5543
25. Park S, Yoon JH, Joo I et al (2022) Image quality in liver CT: low-dose deep learning vs standard-dose model-based iterative reconstructions. Eur Radiol 32:2865–2874
26. Lyu P, Neely B, Solomon J et al (2021) Effect of deep learning image reconstruction in the prediction of resectability of pancreatic cancer: diagnostic performance and reader confidence. Eur J Radiol 141:109825
27. Racine D, Becce F, Viry A et al (2020) Task-based characterization of a deep learning image reconstruction and comparison with filtered back-projection and a partial model-based iterative reconstruction in abdominal CT: a phantom study. Phys Med 76:28–37
28. Nakamura Y, Narita K, Higaki T, Akagi M, Honda Y, Awai K (2021) Diagnostic value of deep learning reconstruction for radiation dose reduction at abdominal ultra-high-resolution CT. Eur Radiol 31:4700–4709
29. Wang X, Zheng F, Xiao R et al (2021) Comparison of image quality and lesion diagnosis in abdominopelvic unenhanced CT between reduced-dose CT using deep learning post-processing and standard-dose CT using iterative reconstruction: a prospective study. Eur J Radiol 139:109735