Deep learning image reconstruction algorithm for abdominal multidetector CT at different tube voltages

assessment of image quality and radiation dose in a phantom study

Park, Hye Joo; Choi, Seo Youn; Lee, Ji Eun; Lim, Sanghyeok; Lee, Min Hee; Yi, Boem Ha; Cha, Jang Gyu; Min, Ji Hye; Lee, Bora; Jung, Yunsub

Published in:
European Radiology

DOI (link to publication from Publisher):
10.1007/s00330-021-08459-8

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

European Radiology (2022) 32:3974–3984
https://doi.org/10.1007/s00330-021-08459-8

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Deep learning image reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of image quality and radiation dose in a phantom study

Hye Joo Park[1] · Seo-Youn Choi[1] · Ji Eun Lee[1] · Sanghyeok Lim[1] · Min Hee Lee[1] · Boem Ha Yi[1] · Jang Gyu Cha[1] · Ji Hye Min[2] · Bora Lee[3,4] · Yunsub Jung[5]

Received: 22 July 2021 / Revised: 15 October 2021 / Accepted: 27 October 2021 / Published online: 22 January 2022

Abstract

Objectives To compare the image quality and radiation dose of a deep learning image reconstruction (DLIR) algorithm compared with iterative reconstruction (IR) and filtered back projection (FBP) at different tube voltages and tube currents.

Materials and methods A customized body phantom was scanned at different tube voltages (120, 100, and 80 kVp) with different tube currents (200, 100, and 60 mA). The CT datasets were reconstructed with FBP, hybrid IR (30% and 50%), and DLIR (low, medium, and high levels). The reference image was set as an image taken with FBP at 120 kVp/200 mA. The image noise, contrast-to-noise ratio (CNR), sharpness, artifacts, and overall image quality were assessed in each scan both qualitatively and quantitatively. The radiation dose was also evaluated with the volume CT dose index ($CTDI_{vol}$) for each dose scan.

Results In qualitative and quantitative analyses, compared with reference images, low-dose CT with DLIR significantly reduced the noise and artifacts and improved the overall image quality, even with decreased sharpness ($p < 0.05$). Despite the reduction of image sharpness, low-dose CT with DLIR could maintain the image quality comparable to routine-dose CT with FBP, especially when using the medium strength level.

Conclusion The new DLIR algorithm reduced noise and artifacts and improved overall image quality, compared to FBP and hybrid IR. Despite reduced image sharpness in CT images of DLIR algorithms, low-dose CT with DLIR seems to have an overall greater potential for dose optimization.

Key Points

• *Using deep learning image reconstruction (DLIR) algorithms, image quality was maintained even with a radiation dose reduced by approximately 70%.*

• *DLIR algorithms yielded lower image noise, higher contrast-to-noise ratios, and higher overall image quality than FBP and hybrid IR, both subjectively and objectively.*

• *DLIR algorithms can provide a better image quality, much better than FBP and even better than hybrid IR, while facilitating a reduction in radiation dose.*

Keywords Multidetector computed tomography · Image reconstruction · Artificial intelligence

✉ Seo-Youn Choi
sychoi@schmc.ac.kr

1 Department of Radiology, Soonchunhyang University Bucheon Hospital, Soonchunhyang University College of Medicine, 170 Jomaru-ro, Bucheon 14584, Republic of Korea

2 Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-Ro Gangnam-gu, Seoul, 06351, Republic of Korea

3 Institue of Public Health and Environment, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

4 Department of Statistics, Chung-Ang University, Seoul, Republic of Korea

5 GE Healthcare Co., Ltd, 416 Hangang-daero, Jung-gu, Seoul, 04637, Republic of Korea

Abbreviations

ASIR	Adaptive statistical iterative reconstruction
CNR	Contrast-to-noise ratio
DCNN	Deep convolutional neural network
DLIR	Deep learning image reconstruction
FBP	Filtered back projection
HU	Hounsfield unit
IR	Iiterative reconstruction
mGy	Milligrays
ROI	Regions of interest
SSIM	Structural similarity

Introduction

Despite the continuous development of CT, medical radiation exposure is still an important issue. Accordingly, in response to these expectations, the dose reduction technique has been constantly developed using tube current modulation, decreasing tube voltage, and image reconstruction algorithms [1–3].

Among them, focusing on the image reconstruction algorithms in more detail, filtered back projection (FBP) was the conventionally used algorithm in image reconstruction owing to its faster reconstruction speed and easy implementation. However, it had several drawbacks in radiation dose reduction in terms of resolution, high image noise, artifacts, and image quality [4]. To overcome the particular shortcomings of FBP, iterative reconstruction (IR) was introduced [5, 6]. Low-dose CT images acquired with the IR technique were evaluated for clinical practice while minimizing image noise and artifacts. Although IR was quite successful in terms of reducing dose and noise simultaneously, which was the main goal, the substantial noise reduction resulted in inevitable "oversmoothing," leading to a "plastic-looking" or "unnatural" appearance of IR-reconstructed images [7]. According to previous studies [8, 9], a smoothed appearance can result in a deterioration of subjective image quality and subsequently a significant decrease in the visibility of small structures, as shown in previous studies. This is why many radiologists were initially inclined to reject the routine implementation of IR algorithms and have questioned the diagnostic acceptability of IR compared to traditional FBP.

Faced with these limitations of IR, deep learning image reconstruction (DLIR) algorithms were developed by GE Healthcare (TrueFidelity™; GE Healthcare), which is trained with high-quality FBP datasets to learn how to differentiate noise from signals [10, 11]. The design goal of DLIR algorithms is to generate a reconstructed image that outperforms previous IR techniques in terms of image quality, dose performance, and reconstruction speed. The DLIR engine generates the output image from an input sonogram that is acquired with low radiation dose, with use of deep convolutional neural networks (DCNNs)-based models. During training, the DCNNs analyze the data and synthesize a reconstruction function, which is optimized through the learning process and extensive testing of dataset for validation. The DLIR is an algorithm developed for image quality similar to "high dose FBP image," as the FBP is the most ideal image reconstruction technique in a high-dose and optimal scan environment. The DLIR technique can be reconstructed in three modes (DLIR-L, M, and H), and the final output image is generated by varying the degree of the noise included in the image for each mode [11–13].

Several studies have attempted to verify the performance of the DLIR technique to date [10, 14–19]; however, to the best of our knowledge, there have only been few studies using the DLIR algorithms developed by GE Healthcare (TrueFidelity™) [10, 17–19] and most of them were conducted using chest CT performed under a single scanning parameter. We thought that a new study is needed to find out how DLIR succeeded in radiation dose reduction while compensating for the weaknesses of IR techniques, especially in various tube voltages and tube currents.

So, the purpose of our study was to compare the image quality, radiation dose, and diagnostic accuracy of a DLIR algorithm (TrueFidelity™) compared with a hybrid IR algorithm (adaptive statistical iterative reconstruction V [ASIR-V]) and FBP at different tube voltages and tube currents.

Materials and methods

Phantom

The trunk and bilateral thigh of the whole-body phantom PBU-60 (Kyoto Kagaku Co., Ltd), simulating a man 165 cm in height and 50 kg in weight, were used in this study (Fig. 1). The phantom was made of a radiological soft tissue substitute with embedded life-sized synthetic skeleton and organs such as the liver with portal and hepatic veins, kidneys, spleen, pancreas, stomach, sigmoid colon, rectum, and prostate. The phantom was placed on the CT bed in a head-first, supine position at the center of the gantry.

Since this study was performed with phantom not human patients, the institutional review board (IRB) approval was not required.

CT acquisition and image reconstruction algorithms

Acquisition of all CT images was performed on a 256-channel multi-detector CT (Revolution, GE Healthcare) equipped with hybrid IR and DLIR algorithms. This scanner can achieve a detector coverage of 160 mm in a single rotation at the isocenter of rotation (detector configuration: 0.625 × 256 mm for

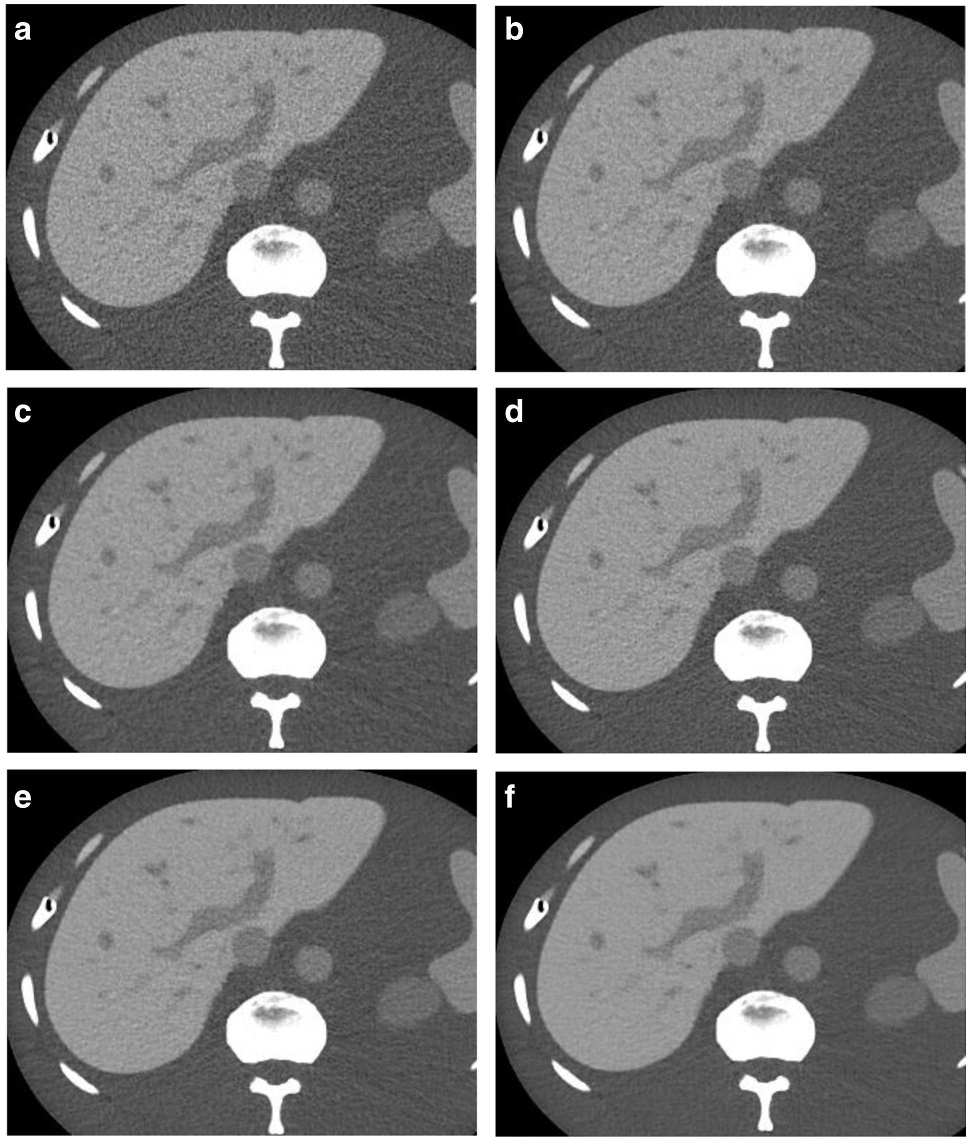

Fig. 1 Axial abdominal CT images acquired with scanning parameters of 80 kVp/200 mA using different reconstruction algorithms: (**a**) filtered back projection (FBP), (**b**) adaptive statistical iterative reconstruction V (ASIR-V) 30%, (**c**) ASIR-V 50%, (**d**) low-level deep learning image reconstruction (DLIR-L), (**e**) medium-level DLIR (DLIR-M), and (**f**) high-level DLIR (DLIR-H). The assessed score of qualitative overall image quality was highest at 4 for images reconstructed with the medium-strength DLIR algorithms. The overall image quality scores were 2, 2, 2, 3, 4, and 3 for the reconstruction algorithms of FBP, ASIR-V 30%, ASIR-V 50%, DLIR-L, DLIR-M, and DLIR-H, respectively

256-row multi-detector CT. The body phantom was scanned with three different tube voltages (80, 100, and 120 kVp) and three different tube currents (60, 100, and 200 mA). According to previous study dealing with image quality depending on reconstruction methods [6], the tube current values of 200 mA and 100 mA were arbitrarily used frequently, and the minimum value of the tube current was set at 60 as it was the minimum tube current allowed in this CT machine (Revolution CT, GE Healthcare). All other CT parameters were kept the same in the nine protocols used (gantry rotation speed of 1 s, pitch of 0.992, and reconstruction slice thickness of 2.5 mm). The scan range was set from the lung bases to the bottom of the pelvic bones. Each of the nine raw research CT data sets was reconstructed on an advanced image processing station (AW4.7, GE Healthcare) with FBP, 30% and 50% ASIR-V®, and three levels of TrueFidelity™ DLIR (low [L], medium [M], and high [H]). The percentage of ASIR-V® was determined as two levels most frequently used in clinical practice and previous studies [6, 10, 12, 13]. Thus, six CT datasets were created for each scan, and ultimately, 54 image series were generated. Basically, we assessed the image quality from different reconstruction algorithms acquired at different dose levels against a well-defined reference standard of FBP at 120 kVp/200 mA.

Radiation dose measurement

CT radiation dose descriptors such as volume CT dose index ($CTDI_{vol}$, described in milligrays [mGy]) were recorded after completion of the CT examination for all image datasets.

Table 1 Grading scale for qualitative analysis of CT examinations

Scale	Image quality parameters			
	Sharpness	Noise	Artifacts	Overall diagnostic acceptability
1	Blurry	Unacceptable noise	Present and affecting image interpretation	Unacceptable
2	Poorer than average	Above-average noise	Present but not affecting interpretation	Suboptimal
3	Average	Average noise	No artifact	Average
4	Better than average	Less-than-average noise	Not applicable	Above average
5	Sharpest	Minimal or no noise	Not applicable	Superior

Assessment of subjective image quality (qualitative analysis)

All randomized CT image datasets were reviewed in a picture archiving and communication system (Deja-View, version 3.0; DongEun IT) by two abdominal radiologists (J.E.L. and S.Y.C., with 9 years and 12 years of experience in abdominal imaging, respectively). As all information such as tube voltage, tube current, and the reconstruction technique used during the CT scan were removed from the image, the reviewers were blinded to such scanning information. For subjective analysis, the two abdominal radiologists scored independently and immediately shared their opinions together. In very few cases of discordance, the final consensus was immediately achieved without delay. During the review, all datasets were displayed at soft tissue settings (window/level, 400/40 Hounsfield unit [HU]) constantly. Evaluated subjective parameters are as follows: sharpness, noise, artifacts, and overall image quality with five- or three-point grading scales [20] (Table 1).

Reviewers first scored each parameter of the images taken at 120 kVp/200 mA with FBP while looking at the criteria, then performed the quality assessments of the other image sets from different reconstructions acquired at different dose levels based on the score of the reference image.

Assessment of objective image quality (quantitative analysis)

Quantitative assessment of noise and contrast-to-noise ratio (CNR) was performed by another abdominal radiologist (S.L., with 15 years of experience in abdominal imaging), who was not involved in the qualitative analysis, using an advanced diagnostic workstation (AW4.7; GE Healthcare). Circular regions of interest (ROI) were drawn in homogenous areas of the right lobe of the liver parenchyma, avoiding large hepatic vessels (400–500 mm^2) and the low-density middle hepatic vein (40–50 mm^2). The sizes and the locations of the ROIs were fixed for all six reconstructed image sets of the same scan parameters. For assessing overall image noise, the standard deviation of the HU was measured three times in homogenous liver parenchyma in our phantom, and the mean value was recorded. Simultaneously, the HUs of the liver parenchyma and middle hepatic vein were also measured three times, and the mean values were recorded. The hepatic vein-to-liver CNR was calculated using the following formula: $\mathrm{CNR} = (\mathrm{ROI_L} - \mathrm{ROI_V})/N$, where $\mathrm{ROI_L}$ is the mean HU of the liver parenchyma, $\mathrm{ROI_V}$ is the mean HU value of the middle hepatic vein, and N is noise. Quantitative measurement of image sharpness and overall image quality was done by a physician (Y.S.J., with 14 years of biomedical engineering and medical image processing/analysis). Image sharpness was assessed using no-reference–based blur metrics, which analyze the sharpness by assessing the behavior of the adjacent pixel variation. The calculated metric value is expressed from 0 to 1, with lower values showing sharper images and higher values indicating blurrier images [17]. To evaluate overall image quality, a structural similarity (SSIM) index was used. SSIM is a metric that assesses structural changes of an image, compared to the reference image (full-reference method), by evaluating elements such as luminance, contrast, and structure (texture information) and calculating the results numerically with a scale ranging from −1 (no similarity) to 1 (identical) [21–24]. The reference image for measuring SSIM was also set as the images taken at 120 kVp/200 mA and reconstructed with the FBP technique.

Statistical analysis

The measurements were analyzed by a nonparametric analysis of covariance (ANCOVA) based on ranks for the dataset comparison with adjustment for the tube voltage and the tube currents. The marginal mean of each measurement was estimated from the linear model adjusted for the tube current and voltage. All the statistical analyses were performed using R (version 4.0.3, The R Foundation for statistical Computing).

Table 2 Scan voltage, current, and radiation dose for each dataset

Current (mA)	Voltage (kV)		
	80	100	120
	CTDIvol	CTDIvol	CTDIvol
60	1.5	2.78	4.35
100	2.5	4.63	7.25
200	4.99	9.25	**14.49**

Units are mGy for CTDIvol. *CTDI* CT dose index. Bold indicates the value of the reference image

Results

Radiation dose

The radiation dose of the scan acquired at 120 kVp/200 mA was 14.49 mGy for $CTDI_{vol}$. The radiation dose decreased by approximately 65% between the 120- and the 80-kVp protocols and by approximately 70% between the 200- and the 60-mA protocols. Radiation doses for each set of scan parameters are summarized in Table 2.

Subjective image quality

First, the reference image was graded as 5, 3, 3, and 4 for sharpness, noise, artifacts, and overall diagnostic acceptability, respectively. Despite a significant reduction in radiation dose, the following low-dose CT with ASIR-V or DLIR images were still graded as having similar or less image noise compared with the reference image: DLIR-H at 80 kVp/60 mA and 100 kVp/60 mA, DLIR-M and DLIR-H at 120 kVp/60 mA and 80 kVp/100 mA, all levels of DLIR at 100 kVp/100 mA, ASIR-V 50% and all levels of DLIR at 80 kVp/100 mA, and all levels of ASIR-V and DLIR at 120 kVp/100 mA and 100 kVp/200 mA. As the radiation dose decreased, the artifacts tended to increase, but ASIR-V rather than FBP and DLIR rather than ASIR-V could complement the artifacts better. As for the sharpness, the highest score of 5 was given for the reference image, and when ASIR-V or DLIR was used other than FBP, and when a higher level of ASIR-V or DLIR was used, the grade lowered. Nevertheless, the overall diagnostic acceptability of the reference image was comparable with those of DLIR images obtained at 120 kVp/100 mA, 100 kVp/200 mA, or 80 kVp/200 mA, especially for medium-level strength. Also, in overall diagnostic acceptability, if the images were taken at the same radiation dose, FBP, ASIR-V, and DLIR were scored in a higher and higher order. Detailed subjective image quality values are summarized in Table 3 (Fig. 2 and Fig. 3).

Objective image quality

Among the quantitative parameters of image quality we assessed, the reference image was defined as the image acquired at 120 kVp/200 mA with the FBP algorithm only for the SSIM index, which was then expressed as a relative value. In contrast, the other three parameters were expressed as absolute values without a reference image. If the radiation dose was equal, the image noise was lowest on DLIR, followed by ASIR-V, and then FBP. Furthermore, as the blending level of ASIR-V and the level of DLIR were increased, the tendency for decreasing image noise was evident. Relatedly, CNR tended to increase more and more with the use of ASIR-V and DLIR, or with higher levels of ASIR-V and DLIR. Interestingly, the blur metrics value was lowest in the images reconstructed with the FBP images, and increased in ASIR-V or DLIR algorithms, indicating that the FBP algorithm resulted in the sharpest images and the ASIR-V and DLIR algorithms resulted in the blurrier images. Additionally, if the DLIR or ASIR-V was used for image reconstruction, the blur metrics increased as the intensity level lowered. The SSIM representing image quality was improved when using a higher radiation dose, higher level of ASIR-V, or higher level of DLIR. In addition, the difference in the SSIM values of images acquired with ASIR-V 50% and DLIR-L was the smallest among the various image sets. The lowest SSIM index was observed for FBP at all dose levels. All of the detailed objective image quality values are displayed in Supplementary Table 1 and Fig. 4 and there was a statistically significant difference between the measured values.

Discussion

In our study, a newly introduced DLIR algorithm yielded lower image noise, higher CNRs, and higher overall image quality, both subjectively and objectively. In addition, in quantitative analysis, the image sharpness of the DLIR images was better than those generated by ASIR-V or even FBP. Thus, we could conclude that the DLIR algorithm can provide a better image quality, much better than FBP and even better than ASIR-V, while reducing radiation dose.

The noise was remarkably reduced in our phantom study when using the DLIR algorithm. The DLIR algorithm TrueFidelity™ was designed to differentiate signal from noise in order to reduce reconstructed image noise without changing its texture [11]. Our study also revealed that DLIR had better noise reduction performance than FBP and ASIR-V, and a progressive reduction in image noise was seen as the DLIR strength increased. In addition, higher CNR was observed with the DLIR algorithm than with FBP and ASIR-V as well, which was related to the noise reduction, and the CNR increased as the DLIR strength increased. Many previous

Table 3 Qualitative analysis of image quality of CT images

Parameters		Sharpness			Noise			Artifact			Overall diagnostic acceptability		
Current	Scan Voltage	80 (kVp)	100	120	80 (kVp)	100	120	80 (kVp)	100	120	80 (kVp)	100	120
60 mA	FBP	2	2	3	1	1	2	1	1	2	1	1	2
	ASIR-V 30%	2	2	3	1	1	2	1	1	2	1	1	2
	ASIR-V 50%	1	1	2	1	2	3	1	2	2	1	1	1
	DLIR-L	2	2	3	1	2	2	1	2	3	1	2	3
	DLIR-M	2	2	2	2	2	3	2	2	3	2	2	3
	DLIR-H	1	1	2	3	3	4	2	3	3	1	1	2
100 mA	FBP	3	3	4	1	2	2	2	1	2	1	1	3
	ASIR-V 30%	3	2	4	1	2	3	2	2	2	1	2	3
	ASIR-V 50%	2	1	3	2	2	3	2	2	3	1	1	2
	DLIR-L	2	2	3	2	3	3	2	2	3	2	2	4
	DLIR-M	2	2	3	3	3	4	3	3	3	2	3	4
	DLIR-H	2	1	2	3	4	5	3	3	3	2	2	3
200 mA	FBP	3	3	**5**	2	2	**3**	3	3	**3**	2	2	**4**
	ASIR-V 30%	3	3	4	2	3	4	3	3	3	2	3	4
	ASIR-V 50%	3	3	3	3	4	5	3	3	3	2	3	4
	DLIR-L	2	2	3	3	4	4	3	3	3	3	4	5
	DLIR-M	2	2	3	3	5	5	3	3	3	4	4	5
	DLIR-H	2	2	2	4	5	5	3	3	3	3	3	4
p value		< 0.0001			< 0.0001			0.0004			< 0.0001		

FBP filtered back projection, *ASIR-V* adaptive statistical iterative reconstruction V, *DLIR-L* low-level deep learning imaging reconstruction, *DLIR-M* medium-level deep learning image reconstruction, *DLIR-H* high-level deep learning imaging reconstruction. Bold indicates the value of the reference image

studies reported that ASIR-V could reduce the noise that is a critical issue in dose reduction, and this advantage of ASIR-V has been acknowledged until now [3, 6, 20, 25, 26]. Interestingly, in our study, although the objectively assessed noise in the images with middle- or high-level DLIR was lower than that obtained with ASIR-V 50%, the subjectively assessed noise was lower even in the images with low-level DLIR as compared to those with ASIR-V 50%. This slight difference might be because the quantitative value of noise, which is evaluated only by the SI of HU, has limitations in accurately representing other characteristics such as texture [27]. However, it should be noted that the image noise of the DLIR algorithm has a certain improving tendency over FBP and even over ASIR-V.

Blur metrics was initially proposed in 1997 and has been applied to control and quantify the image sharpness by assessing the blur effect of the edge [28–31]. The trade-off relationship between noise and sharpness is one of the weaknesses of IR-based image reconstruction methods [7, 20]. As the number of iterations increases, the noise disappears, but the image sharpness also decreases. In quantitative and quantitative analyses using blur metrics in our study, the image sharpness degraded when using not only DLIR algorithm rather than FBP but also higher levels of DLIR. The degradation of image sharpness resulting in a blotchy appearance of the solid organs could compromise the diagnostic accuracy if it caused a well-defined lesion to appear less well-defined. Such degradation can make it difficult to differentiate a small hepatic cyst from metastasis in abdominal CT interpretation for radiologists and had been regarded as an important drawback of ASIR [32–34]. At the beginning of this study, we expected that DLIR has a great advantage in improving the noise and sharpness simultaneously, which was regarded to be the weakest point of ASIR-V in terms of image sharpness. In addition, there was a previous study by Nakamura et al. [35] which reported that DLIR improved the quality of abdominal CT images for the evaluation of hypovascular hepatic metastases, as compared with hybrid IR. However, it should be noted that in the subjective and objective analyses in our study, the images became more blurred in either ASIR-V or DLIR than in FBP. As it is not possible to compare the blending percentage of ASIR-V and the mode of DLIR in one-to-one matching, it would be difficult to say which of the two, ASIR-V and DLIR, has more weakness in terms of the sharpness. However, despite these tough comparisons, considering other evaluated parameters

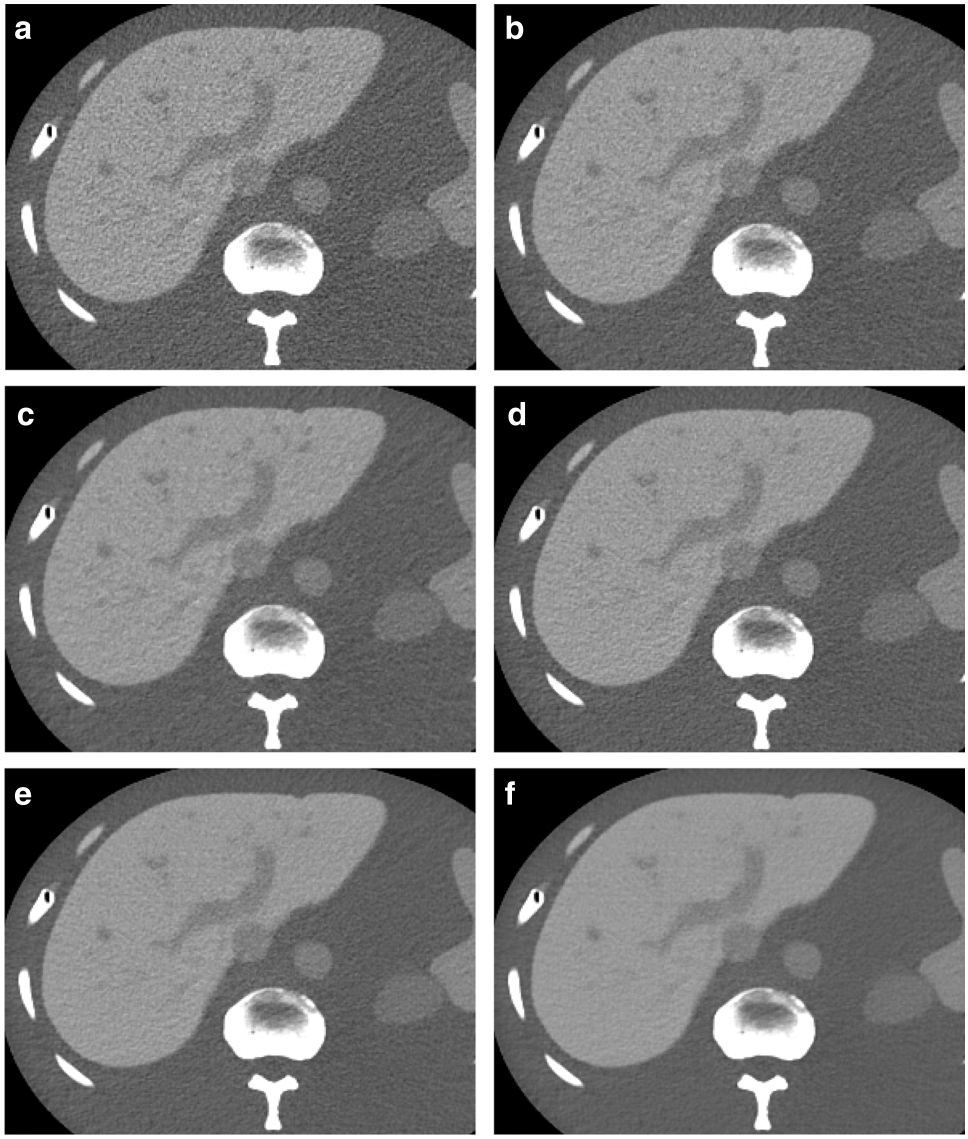

Fig. 2 Axial abdominal CT images acquired with scanning parameters of 100 kVp/100 mA using different reconstruction algorithms: **a** filtered back projection (FBP), **b** adaptive statistical iterative reconstruction V (ASIR-V) 30%, **c** ASIR-V 50%, **d** low-level deep learning image reconstruction (DLIR-L), **e** medium-level DLIR (DLIR-M), and **f** high-level DLIR (DLIR-H). The assessed score of qualitative overall image quality was highest at 4 for images reconstructed with the medium-strength DLIR algorithms. The overall image quality scores were 2, 2, 2, 3, 4, and 3 for the reconstruction algorithms of FBP, ASIR-V 30%, ASIR-V 50%, DLIR-L, DLIR-M, and DLIR-H, respectively

especially subjective overall diagnostic acceptability and objective SSIM index, we expect DLIR to be a better alternative in real clinical practice. We assume that there might be interfering factors due to other parameters that are inseparable in the visual assessment such as noise, texture, and artifacts. To the best of our knowledge, although there have been few attempts to assess the CT image sharpness objectively so far [36–38], further studies are needed to apply these values to actual clinical imaging.

Several studies have evaluated the image quality of CT with DLIR algorithm by assessing various indicators including the noise, CNR, artifacts, spatial resolution, and diagnostic acceptability of CT images [10, 12–15, 39]. Although a few studies tried to assess the overall image quality quantitatively [22, 40, 41], to the best of our knowledge, there has been no study evaluating the overall image quality represented by objective indices in terms of abdominal CT or DLIR algorithm yet. In our study, we focused on the clinically relevant application of SSIM, which is a numerical indicator for perceived image quality, in clinically available CT reconstruction techniques. Interestingly, all of the reported SSIM indices in our study were relatively high (> 0.99 on a scale from 0 to 1) even for low-dose scanning. Although it was a very small absolute difference, there was a certain directionality to the SSIM index, and considering the previous literature [22, 41], such a small difference in SSIM index might be regarded as a clinically relevant difference. Based on the overall image quality results in our study, when DLIR is used, the tube current should be set at 200 mA to lower the scan voltage to 100kVp or less, and the scan voltage should be set at 120kVp to lower the tube current to 100 mA, to obtain images of considerable quality compared to that of a full-dose scan. However, since our study did not evaluate the tube current and scan voltage at tight

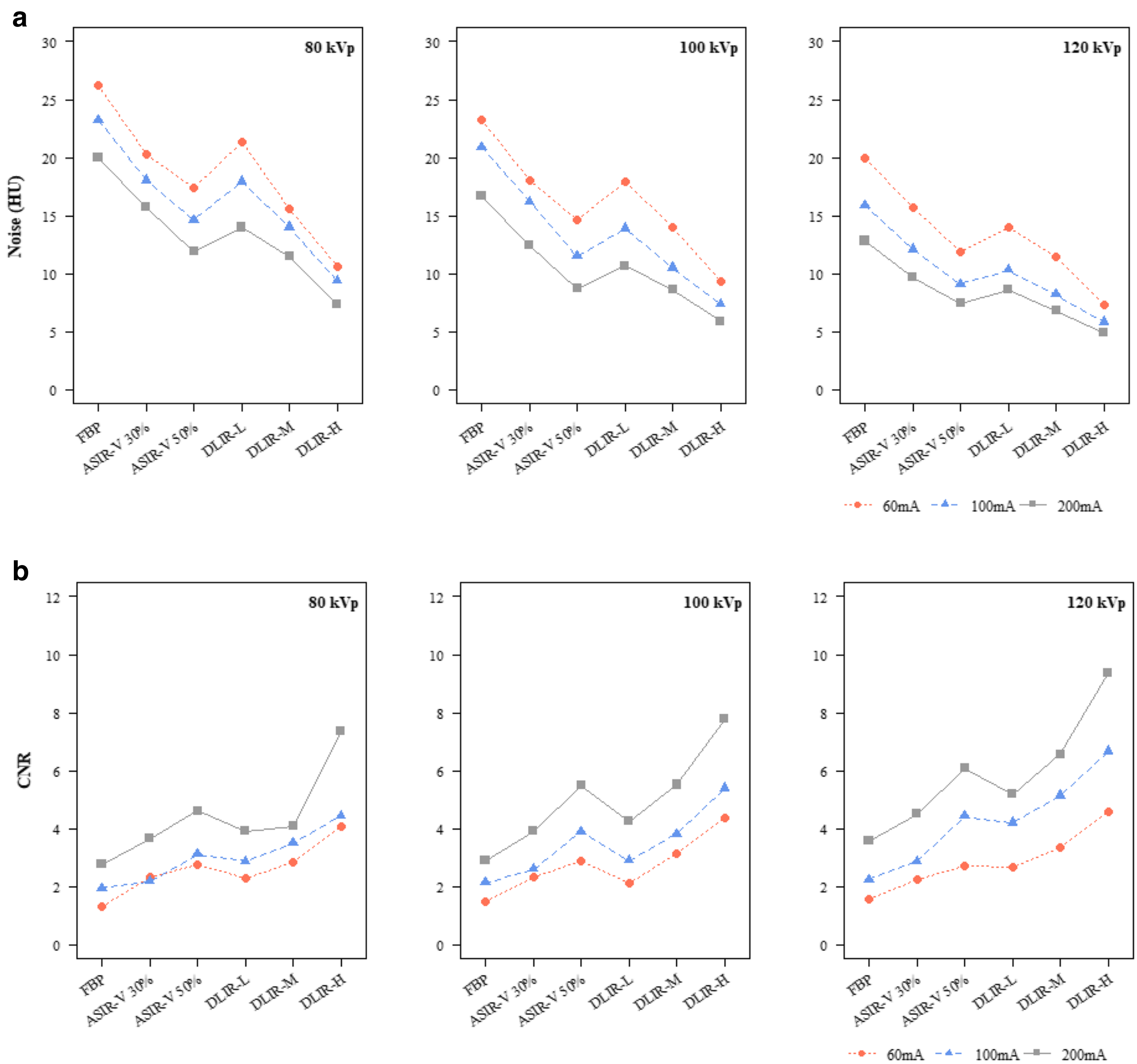

Fig. 3 Line graphs representing objective image quality according to different image reconstruction algorithms with different radiation doses. When the deep learning image reconstruction (DLIR) was used for image reconstruction and as the DLIR strength increases, (**a**) the noise (measured CT number standard deviation) illustrated a progressive reduction, (**b**) the contrast-to-noise ratio was progressively increased, (**c**) the sharpness (measured by blur metrics) was improved when using DLIR rather than filtered back projection (FBP), but progressively worsened as the DLIR strength increased, and (**d**) the overall image quality (represented by structural similarity) was progressively improved. *ASIR-V* adaptive statistical iterative reconstruction V

intervals, further study is needed. Furthermore, considering all evaluated parameters, medium strength is considered to produce images of best quality among the three different DLIR strength levels.

Our study has several potential limitations. First, our body phantom was simplified anatomically to mimic a real human body, and there might be a significant difference from actual body imaging, in which most are enhanced with contrast administration. Further human studies of abdominal CT with contrast enhancement will be needed in the future. Second, our study used a single scanner, so the question remains whether our results will be implemented in the same way on other scanners. Third, we did not perform an analysis of the contrast and spatial resolution, which should be done. Fourth, we had only one body phantom that mimicked a real human body and could not perform the same evaluation by

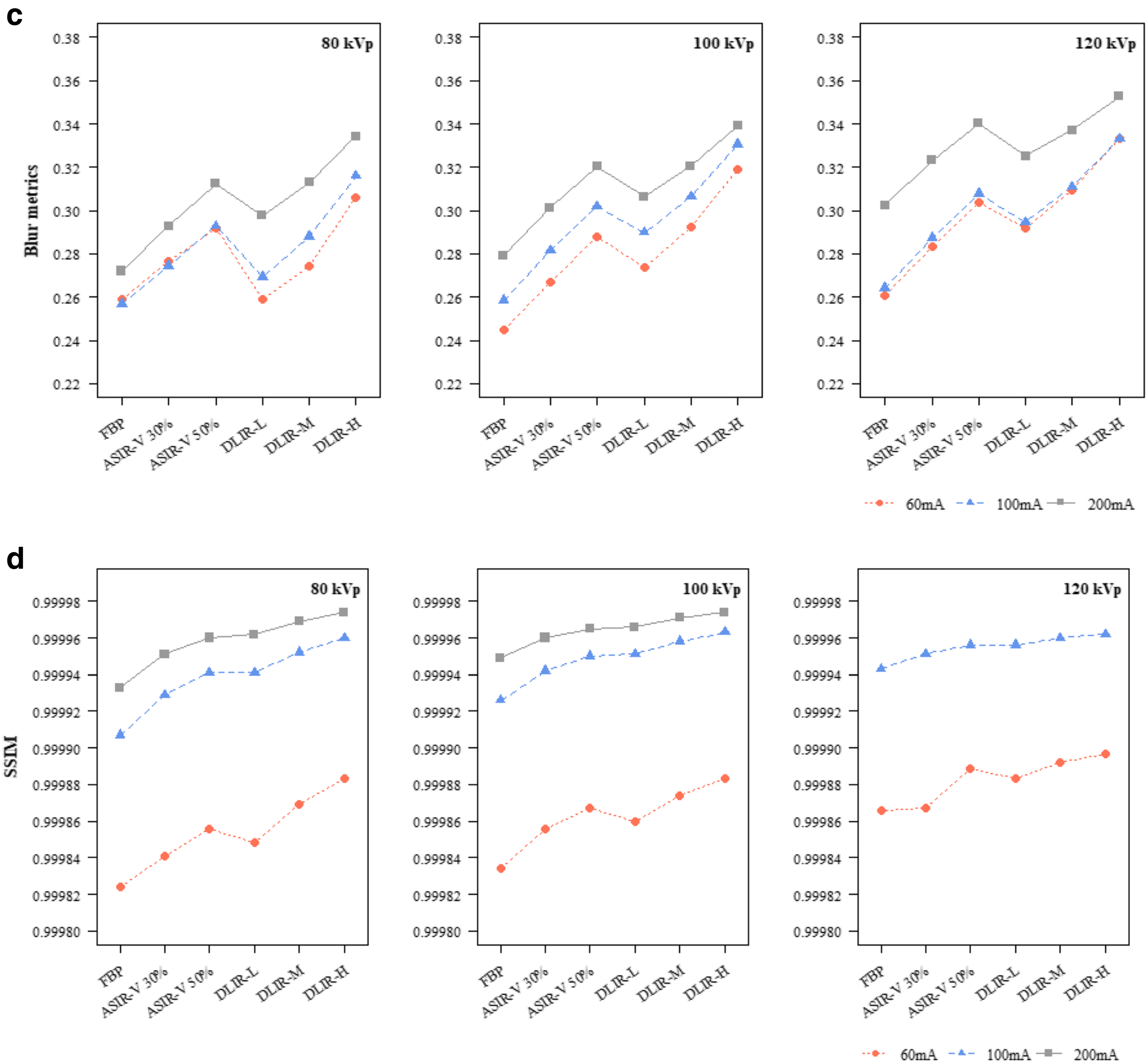

Fig. 3 (continued)

using a different-sized phantom or real human body. Fifth, a bias may have been involved because the observers reviewed the reference standard image first, and then reviewed the quality of other images based on this. Finally, since blur metrics have not been evaluated in radiologic images so far, it is necessary to verify how well they are applied in actual clinical images, rather than for engineering.

In conclusion, new DLIR algorithms reduced noise and artifacts and improved overall image quality, compared to FBP and hybrid IR. Despite reduced image sharpness in CT images of DLIR algorithms, low-dose CT with DLIR seems to have an overall greater potential for dose optimization.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-021-08459-8.

Acknowledgements This work was supported by the Soonchunhyang University Research Fund. The authors thank Kyoung-A Um for their technical support.

Funding This work was supported by the Soonchunhyang University Research Fund.

Declarations

Guarantor The scientific guarantor of this publication is Seo-Youn Choi, M.D. PhD..

Conflict of interest One of the authors (Yunsub Jung) is an employee of GE Healthcare. The remaining authors declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors is a statistician (Bora Lee, PhD, of medical statistics with 15 years of experience and majored in statistics) kindly provided statistical advice for this manuscript.

Informed consent This study was performed with a phantom not human patients or animal; institutional review board (IRB) approval was not required and informed consent was not needed as well.

Ethical approval This study was performed with a phantom not human patients or animal; institutional review board (IRB) approval was not required and informed consent was not needed as well.

Methodology • experimental

• performed at one institution

References

1. Kaza RK, Platt JF, Goodsitt MM et al (2014) Emerging techniques for dose optimization in abdominal CT. Radiographics 34:4–17
2. Nakayama Y, Awai K, Funama Y et al (2005) Abdominal CT with low tube voltage: preliminary observations about radiation dose, contrast enhancement, image quality, and noise. Radiology 237:945–951
3. Schindera ST, Diedrichsen L, Müller HC et al (2011) Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study. Radiology 260:454–462
4. Geyer LL, Schoepf UJ, Meinel FG et al (2015) State of the art: iterative CT reconstruction techniques. Radiology 276:339–357
5. Silva AC, Lawder HJ, Hara A et al (2010) Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. AJR Am J Roentgenol 194:191–199
6. Singh S, Kalra MK, Hsieh J et al (2010) Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. Radiology 257:373–383
7. Leipsic J, Labounty TM, Heilbron B et al (2010) Adaptive statistical iterative reconstruction: assessment of image noise and image quality in coronary CT angiography. AJR Am J Roentgenol 195:649–654
8. Böning G, Schäfer M, Grupp U et al (2015) Comparison of applied dose and image quality in staging CT of neuroendocrine tumor patients using standard filtered back projection and adaptive statistical iterative reconstruction. Eur J Radiol 84:1601–1607
9. Vachha B, Brodoefel H, Wilcox C et al (2013) Radiation dose reduction in soft tissue neck CT using adaptive statistical iterative reconstruction (ASIR). Eur J Radiol 82:2222–2226
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(7):3951–3959
11. Hsieh J, Liu E, Nett B et al (2019) A new era of image reconstruction: TrueFidelity^{TM}: technical white paper on deep learning image reconstruction. Available via https://www.gehealthcare.ru/-/jssmedia/040dd213fa89463287155151fdb01922.pdf. Accessed May 1 2020
12. 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(1):50–57
13. Singh R, Digumarthy SR, Muse VV et al (2020) Image quality and lesion detection on deep learning reconstruction and iterative reconstruction of submillisievert chest and abdominal CT. AJR Am J Roentgenol 214:566–573
14. Akagi M, Nakamura Y, Higaki T et al (2019) Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. Eur Radiol 29:6163–6171
15. Tatsugami F, Higaki T, Nakamura Y et al (2019) Deep learning-based image restoration algorithm for coronary CT angiography. Eur Radiol 29:5322–5329
16. Shin YJ, Chang W, Ye JC et al (2020) Low-dose abdominal CT using a deep learning-based denoising algorithm: a comparison with CT reconstructed with filtered back projection or iterative reconstruction algorithm. Korean J Radiol 21:356–364
17. Park C, Choo KS, Jung Y et al (2020) CT iterative vs deep learning reconstruction: comparison of noise and sharpness. Eur Radiol. https://doi.org/10.1007/s00330-020-07358-8
18. Kim JH, Yoon HJ, Lee E et al (2021) Validation of deep-learning image reconstruction for low-dose chest computed tomography scan: emphasis on image quality and noise. Korean J Radiol 22:131–138
19. Hata A, Yanagawa M, Yoshida Y et al (2021) The image quality of deep-learning image reconstruction of chest CT images on a mediastinal window setting. Clin Radiol 76:155.e115-155.e123
20. Sagara Y, Hara AK, Pavlicek W et al (2010) Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients. AJR Am J Roentgenol 195:713–719
21. Wang Z, Bovik AC, Sheikh HR et al (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13:600–612
22. Joemai RMS, Geleijns J (2017) Assessment of structural similarity in CT using filtered backprojection and iterative reconstruction: a phantom study with 3D printed lung vessels. Br J Radiol 90:20160519
23. Peng J, Shi C, Laugeman E et al (2020) Implementation of the structural SIMilarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection. Med Phys 47:1907–1919
24. Renieblas GP, Nogués AT, González AM et al (2017) Structural similarity index family for image quality assessment in radiological images. J Med Imaging (Bellingham) 4:035501
25. Pauchard B, Higashigaito K, Lamri-Senouci A et al (2017) Iterative reconstructions in reduced-dose CT: which type ensures diagnostic image quality in young oncology patients? Acad Radiol 24:1114–1124
26. Greffier J, Macri F, Larbi A et al (2016) Dose reduction with iterative reconstruction in multi-detector CT: what is the impact on deformation of circular structures in phantom study? Diagn Interv Imaging 97:187–196
27. Al-Kadi OS (2010) Assessment of texture measures susceptibility to noise in conventional and contrast enhanced computed tomography lung tumour images. Comput Med Imaging Graph 34:494–503
28. Chiang MC, Boult TE (1997) Local blur estimation and super-resolution. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, San Juan, PR 821–826
29. Marichal X, Ma WY, Zhang HJ (1999) Blur determination in the compressed domain using DCT information. Proceedings of the IEEE International Conference on Image Processing. IEEE, Kobe, JP 386–390

30. Marziliano P, Dufaux F, Winkler S et al (2002) A no-reference perceptual blur metric. Proceedings of the IEEE International Conference on Image Processing. IEEE, Rochester, NY 57–60
31. Caviedes J, Gurbuz S (2002) No-reference sharpness metric based on local edge kurtosis. Proceedings of the IEEE International Conference on Image Processing. IEEE, Rochester, NY 53–56
32. Feger S, Kendziorra C, Lukas S et al (2018) Effect of iterative reconstruction and temporal averaging on contour sharpness in dynamic myocardial CT perfusion: sub-analysis of the prospective 4D CT perfusion pilot study. PLoS One 13:e0205922
33. Shieh CC, Kipritidis J, O'Brien RT et al (2014) Image quality in thoracic 4D cone-beam CT: a sensitivity analysis of respiratory signal, binning method, reconstruction algorithm, and projection angular spacing. Med Phys 41:041912
34. Harris MA, Huckle J, Anthony D et al (2017) The acceptability of iterative reconstruction algorithms in head CT: an assessment of sinogram affirmed iterative reconstruction (SAFIRE) vs. filtered back projection (FBP) using phantoms. J Med Imaging Radiat Sci 48:259–269
35. Nakamura Y, Higaki T, Tatsugami F et al (2019) Deep learning–based CT image reconstruction: initial evaluation targeting hypovascular hepatic metastases. Radiol Artif Intell 1:e180011
36. Verdun FR, Racine D, Ott JG et al (2015) Image quality in CT: from physical measurements to model observers. Phys Med 31:823–843
37. Eldevik K, Nordhøy W, Skretting A (2010) Relationship between sharpness and noise in CT images reconstructed with different kernels. Radiat Prot Dosimetry 139:430–433
38. Kayugawa A, Ohkubo M, Wada S (2013) Accurate determination of CT point-spread-function with high precision. J Appl Clin Med Phys 14:3905
39. Higaki T, Nakamura Y, Zhou J et al (2020) Deep learning reconstruction at CT: phantom study of the image characteristics. Acad Radiol 27:82–87
40. Wu D, Kim K, Li Q (2019) Computationally efficient deep neural network for computed tomography image reconstruction. Med Phys 46:4763–4776
41. Buty M, Xu Z, Wu A et al (2017) Quantitative image quality comparison of reduced- and standard-dose dual-energy multiphase chest, abdomen, and pelvis CT. Tomography 3:114–122