

Image Quality Assessment of Deep Learning Image Reconstruction in Torso Computed Tomography Using Tube Current Modulation

Kazuhiro Takeuchi^{a,b}, Yasuhiro Ide^a, Yuichiro Mori^a, Yusuke Uehara^a,
Hiroshi Sukeishi^a, and Sachiko Goto^{b*}

^aDepartment of Radiology, Kagawa University Hospital, Miki, Kagawa 761-0793, Japan,

^bDepartment of Radiological Technology, Graduate School of Health Sciences, Okayama University, Okayama 700-8558, Japan

Novel deep learning image reconstruction (DLIR) reportedly changes the image quality characteristics based on object contrast and image noise. In clinical practice, computed tomography image noise is usually controlled by tube current modulation (TCM) to accommodate changes in object size. This study aimed to evaluate the image quality characteristics of DLIR for different object sizes when the in-plane noise was controlled by TCM. Images acquisition was performed on a GE Revolution CT system to investigate the impact of the DLIR algorithm compared to the standard reconstructions of filtered-back projection (FBP) and hybrid iterative reconstruction (hybrid-IR). The image quality assessment was performed using phantom images, and an observer study was conducted using clinical cases. The image quality assessment confirmed the excellent noise-reduction performance of DLIR, despite variations due to phantom size. Similarly, in the observer study, DLIR received high evaluations regardless of the body parts imaged. We evaluated a novel DLIR algorithm by replicating clinical behaviors. Consequently, DLIR exhibited higher image quality than those of FBP and hybrid-IR in both phantom and observer studies, albeit the value depended on the reconstruction strength, and proved itself capable of providing stable image quality in clinical use.

Key words: computed tomography, deep learning, image reconstruction, tube current modulation, object size

Computed tomography (CT) examinations are an essential and universal tool in medical institutions because of their high diagnostic imaging ability, high throughput, and few restrictions. As the number and type of CT examinations increase, the share of CT in medical radiographic exposure is also increasing, making the optimization of CT dose an essential consideration [1,2]. Filtered-back projection (FBP), which is the basic image reconstruction algorithm of CT, has a linear relationship between dose and image quality, while the usual relationship between image quality and CT dose is a trade-off, as it becomes difficult to reduce image noise at lower dose levels. Iterative reconstruction (IR) have been reported, and modalities such as hybrid-IR and model-based IR (MBIR) are used in clinical practice to maintain image quality while reducing image noise [3,4].

However, several problems have been reported with such IR algorithms, and the noise level of the image and reconstruction strengths are both known to affect the image texture [5]. Thus, it is difficult to reduce the dose significantly below the levels made possible by FBP [6,7].

Conversely, deep-learning image reconstruction (DLIR), which was newly developed by the deep neural network (DNN), reportedly reduces noise and dose without affecting the texture of the image [8-12,13].

Received May 22, 2022; accepted September 1, 2022.

*Corresponding author. Phone: +81-86-235-6907; Fax: +81-86-222-3717
E-mail: goto@md.okayama-u.ac.jp (S. Goto)

Conflict of Interest Disclosures: No potential conflict of interest relevant to this article was reported.

This novel DLIR is a nonlinear image reconstruction algorithm that depends on image noise and contrast and is affected by the subject size in clinical practice. In clinical applications, tube current modulation (TCM) is applied to optimize the incident dose according to the subject size [14]. TCM is a radiation exposure reduction technique that optimizes the incident dose for each patient by assessing the axial changes in body thickness from scout images and regulating the X-ray dose by modulating the tube current to fit the irregular patient contours.

Solomon *et al.* [15] used the Mercury Phantom V3.0 to evaluate the image quality of DLIR and sampled different object sizes using a fixed tube current, which was necessary to maintain measurement accuracy. However, the Mercury phantom is useful only for the clinical evaluation of TCM [16]. Thus, in this study, we assumed a clinical CT of the torso, and the task of image quality assessment targeted two inserts in the Mercury Phantom of low-to-medium contrast: polystyrene ($\approx \Delta 50$ HU) and water ($\approx \Delta 90$ HU). For advanced reconstruction methods such as DLIR, high-contrast objects may be easily distinguished from noise and edges based on previous physical and clinical evaluations [8, 11].

There have been few reports on the physical evaluation of DLIR related to the measurement of low-contrast objects of 50 HU. Furthermore, image quality characteristics using IR at low contrast show different characteristics according to the apparatus [17]. A 50-HU contrast for CT in the abdominal region is clinically important for detecting faint shadows such as those created by liver tumors and intravenous thrombi [18-20]. Thus, this study aimed to physically and visually evaluate the image quality characteristics of different subject sizes reconstructed by DLIR using images acquired under controlled in-plane noise using TCM.

Materials and Methods

Deep-learning image reconstruction. Developed by GE, DLIR (true fidelity imaging [GE Healthcare, Waukesha, WI, USA]) is a DNN-based reconstructive algorithm that uses high-quality FBP data for teacher data in the learning process. DLIR has low, medium, and high setting strengths, each reconstructed from a different DNN model. The amount of noise reduction in the image depends on the strength selected. The tra-

ditionally used hybrid-IR (ASiR-V; GE Healthcare, Waukesha, WI, USA) adjusts the strength by blending FBP and IR [9].

Object-size modulation phantom. A Mercury 4.0 Phantom (Gammex, Middleton, WI, USA) was developed to evaluate TCM installed in CT systems [21, 22]. This phantom was approved by the American Association of Physicists in Medicine (AAPM) Task Group (TG) 233 [22]. The Mercury 4.0 Phantom has five sections with different diameters (16, 21, 26, 31, and 36 cm) made of polyethylene (Fig. 1). The phantom has a uniform layer for evaluating noise characteristics, along with five cylinders of water, bone, polystyrene, 10 mg/mL iodine, and air at a constant distance from the center as contrast inserts for evaluating the resolution. By scanning this phantom with TCM, it is possible to acquire images under clinical conditions, and comprehensive image quality assessment can be performed by evaluating the image quality at each phantom size.

Data acquisition and image reconstruction. Data in this study were acquired using a 256-slice Revolution CT (GE Healthcare, Waukesha, WI, USA) with hybrid-IR and DLIR as image-reconstruction algorithms. The X-ray tube voltage was 120 kV, whereas the TCM determined the X-ray tube current, which was controlled using a noise index (NI). The current was modulated at 10-720 mA using a phantom diameter and

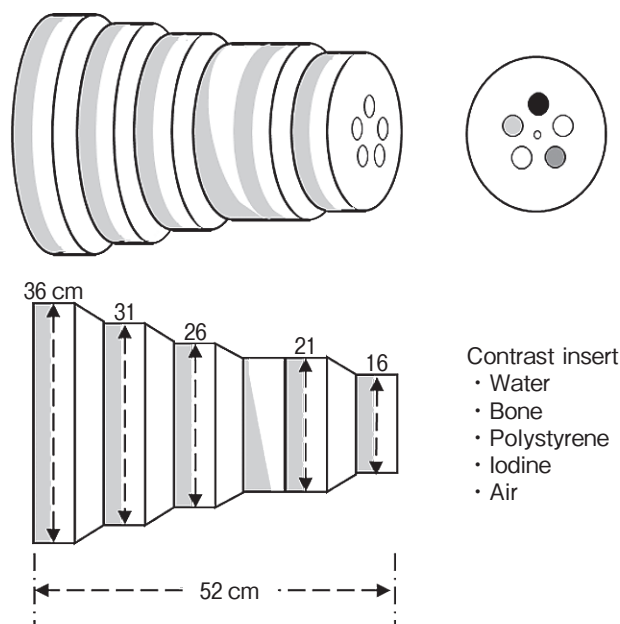


Fig. 1 Composition of the Mercury Phantom used in this study.

NI. NI is determined by the operator, and the imaging dose is controlled to have a constant standard deviation (SD) in the central region of the phantom, regardless of patient size [23]. Thus, the NI is approximately equivalent to the SD of the central region reconstructed with the standard kernel of the FBP using a uniform phantom. In this study, two NIs (13.7 and 22.4) were set, in which the tube current varied without saturation at each phantom size within the range of the upper and lower limits of the TCM (10-720 mA). Considering the guidance level of the adult abdomen announced by the International Atomic Energy Agency, NI=13.7 is the standard dose level, and NI=22.4 is the low dose level [24]. The volume CT dose index (CTDIvol) reported by the scanner console was recorded in a DICOM radiation dose structured report file after each scan. The mean CTDIvol values recorded from the apparatus were 11.6 mGy at the standard dose level and 3.07 mGy at the low dose level. Nominal CTDIvols with phantom diameters of 16, 21, 26, 31, and 36 cm were 2.0, 3.3, 7.4, 17.5, and 28.8 mGy, at standard doses and 0.73, 1.18, 2.45, 4.81, and 6.49 mGy at low doses, respectively. The detector configuration was 0.625 mm × 256 rows, and the rotation time was 0.6 s/rotation. When NI=13.7, the focal spot size was XL, and when NI=22.4, the focal spot size was S. The images were then subjected to FBP, hybrid IR, and DLIR. Hybrid-IR was evaluated for the ASiR-V50 (IR50; a 50 × 50 combination of FBP and IR), which is frequently used in clinical practice, and IR100, which shows strong performance. DLIR was evaluated for three strengths: low (DL-L), medium (DL-M), and high (DL-H). Images were reconstructed with a standard kernel: slice thickness, 1.25 mm (gap less); field of view, 400 mm; and matrix size, 512 × 512 pixels.

Image quality assessment. ImQuest, developed by the clinical imaging physics group at Duke University, was used to analyze the acquired images. ImQuest is an open-source software that uses the technology described in TG233 of the AAPM and is compatible with the Mercury 4.0 Phantom. To assess the image quality characteristics, the noise power spectrum (NPS), task-based modulation transfer function (TTF), and task-based detectability index (d') were calculated using ImQuest. The quality and quantity of the in-plane noise were evaluated using the NPS. As shown in Fig. 2A, we established regions of interests (ROIs) of 40 × 40 pixels and obtained NPS from 60 consecutive axial slices (240 ROIs) in each section with uniform

layers [22]. The calculated NPS maximum was defined as the NPS_{peak} , and the average of the obtained NPS values was quantified as the $NPS_{average}$ spatial frequency. As shown in Fig. 2B, the resolution was measured for two low-to-medium contrast inserts: polystyrene and water. A circular ROI was set for each insert, and the TTF was evaluated using the circular edge technique by taking the average of 80 consecutive axial slices in each section [22, 25, and 26]. The spatial frequency that was 0.5 of the normalized TTF value was quantified as TTF50%. Conversely, d' is an index that can be virtually simulated as the detectability by a radiologist using the measured NPS and TTF and by setting the detection task that is clinically required using the following equation:

$$d'^2 = \frac{\iint |W(u, v)|^2 \cdot TTF(u, v)^2 \cdot E(u, v)^2 \, dudv}{\iint |W(u, v)|^2 \cdot TTF(u, v)^2 \cdot NPS(u, v)^2 \cdot E(u, v)^4 \, dudv}$$

where u and v are the spatial frequencies of the x- and y-coordinates, respectively. $NPS(u, v)$ and $TTF(u, v)$ are measured from the phantom, and values reflecting the tumor size to be simulated are used. $W(u, v)$ is a clinical task function assuming various arbitrary tumor sizes. Conversely, $E(u, v)$ is an eye filter that models the sensitivity of the human visual system to various spatial frequencies and requires the setting of observa-

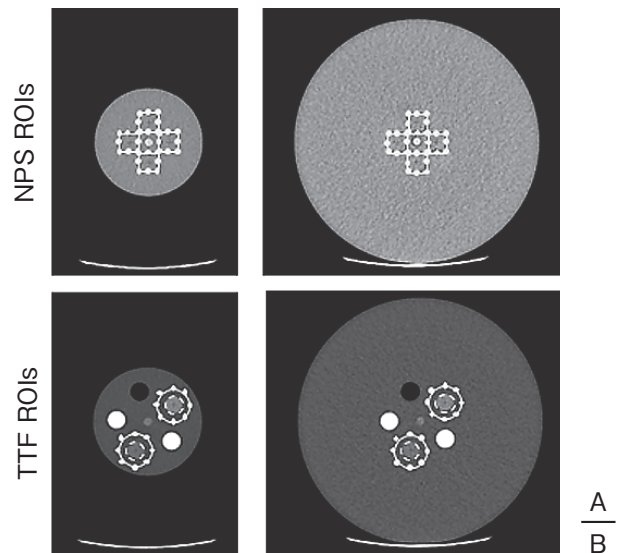


Fig. 2 A, Locations of the regions of interest for the noise power spectrum measurements; B, ROIs located for task-based modulation transfer function measurements with polystyrene and water inserts.

tion conditions using the following equation [27]:

$$E(u, v) = \left| \eta \cdot \sqrt{u^2 + v^2} \cdot \left(\frac{FOV \cdot R \cdot \pi}{D \cdot 180} \right)^{\alpha_1} \cdot \exp \left\{ -\alpha_2 \cdot \left(\sqrt{u^2 + v^2} \cdot \frac{FOV \cdot R \cdot \pi}{D \cdot 180} \right)^{\alpha_3} \right\} \right|^2$$

where η normalizes the function so that its maximum value is one. α_1 , α_2 , and α_3 are the constant parameters 1.5, 3.22, and 0.68, respectively. FOV is the reconstructed field of view, R is the viewing distance, and D is the display size. The calculation condition of d' was based on a non-prewhitening matched filter with an eye filter (NPWE), a zoom factor of 1.5, and a simulated tumor expressed in a Gaussian field of 8 mm [18, 22, 27, and 28]. In this study, we simulated target tumors in the abdominal region and used two low-to-medium-contrast TTFs, NPS values obtained from different doses, and phantom sizes. Subsequently, the rate of increase ($d'\%$) of each algorithm for FBP was calculated from the obtained d' .

Observer study. Five clinical CT scans were randomly selected from the cases obtained for medical purposes in the observer study. Two regions were selected from the five cases acquired using TMC. The neck was selected as a body part where the subject size varied greatly, and the abdomen was selected as a part where the subject size in the torso was large, making low-contrast visibility important. For the paired comparison method, two images from four different reconstruction methods were selected for one CT dataset (12 combinations), and this was conducted on 10 datasets (five cases with two regions). Fig. 3 shows a sample of

the patient used in the observer study. Since there was a possibility that the image noise would vary depending on the subject size even when the TMC was used, SD was measured for the neck and abdomen in each case by placing ROIs in the soft tissue areas. Five radiological technologists with more than 10 years of experience in reading CT images participated in the observation. Previously an observer study using Ura's method of Scheffé's paired comparison was conducted using the software developed by Shiraishi *et al.* [29, 30], based on the ROCKIT developed by Metz *et al.* [31] The reconstruction conditions (FBP, DL-H, IR50, IR100) were varied and compared for each case, as shown in Fig. 4. The window width and window level of the CT images were fixed, while viewer functions such as zoom were restricted. The two images displayed to the observer for

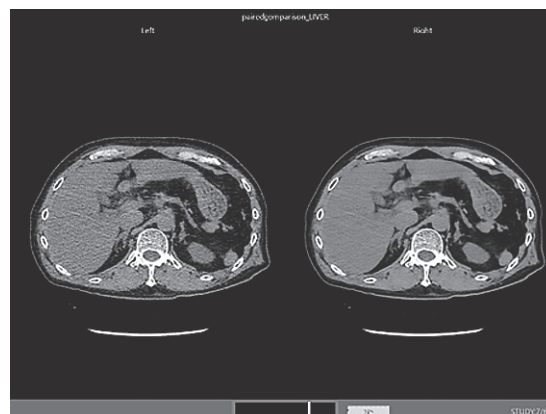


Fig. 4 Display of ROC Viewer to compare computed tomography images using different reconstruction methods.

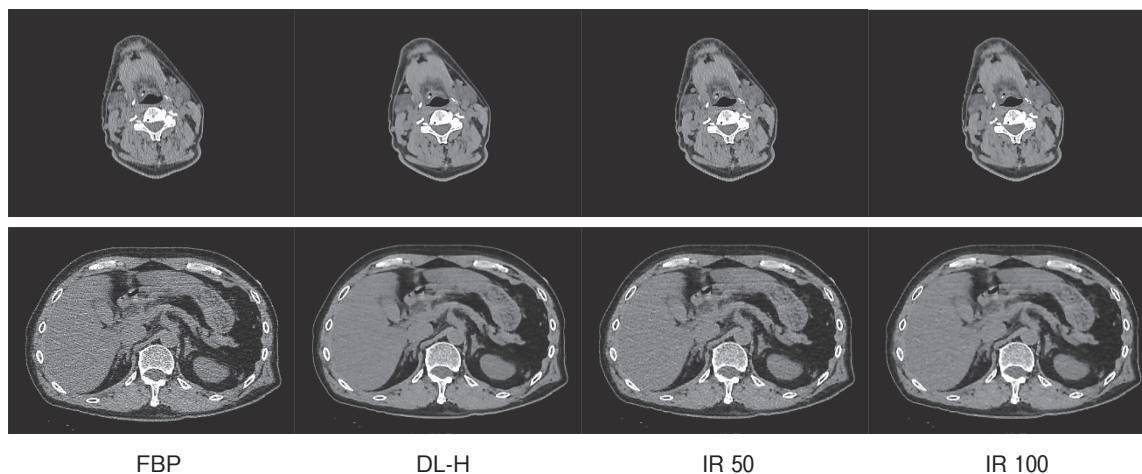


Fig. 3 Examples of case samples for the paired comparison observer study. Four computed tomography (CT) images with different reconstruction techniques were created from CT images of the neck and abdomen.

each body part were rated comprehensively for noise, contrast, and image texture. The results of each combination were analyzed using an ROC-analyzer, and the yardstick was calculated using the multiple comparison method. A statistically significant difference was considered to exist when the difference in average psychological measures between the samples was greater than the yardstick ($p < 0.05$).

Ethical considerations. This study is based on anonymized patient data and is not likely to affect patients. Approval for use of their data in this study was obtained from the Ethical Committee of the Kagawa University School of Medicine (2021-163). Participation in the observer study was thoroughly explained and consent obtained.

Results

Fig. 5 shows a portion of the NPS results at low doses, and Table 1 shows the results of the NPS_{peak} and $NPS_{average}$ at each dose level and phantom size. NPS_{peak} showed similar values depending on the TCM with the

set NIs; however, the values were smallest when the phantom size was 16 cm, and largest when the phantom size was 36 cm, and fluctuated slightly at the 21, 26, and 31 cm diameters. This trend was also observed for different doses. In all conditions, FBP had the highest noise level, and DL-H had the lowest noise level. The NPS average frequency of DLIR compared with FBP was almost similar (0.01 - 0.04 mm^{-1} at a low dose and 0 - 0.03 mm^{-1} at a normal dose), and the difference between FBP and IR50 was approximately 0.04 mm^{-1} , regardless of dose. The difference between FBP and IR50 was approximately 0.04 mm^{-1} regardless of dose, but the difference between FBP and IR100 was very large (approximately 0.18 mm^{-1} at the low dose and approximately 0.17 mm^{-1} at the normal dose). Phantom size had no specific effect on NPS.

Fig. 6 shows some of the TTF results calculated for polystyrene and water at low doses, while Table 2 shows the TTF50% results for each dose level and phantom size. The average deviation for each dose and phantom size was 4.2%. For TTF polystyrene, FBP was the highest, while DL-M and IR50 were similar. Compared

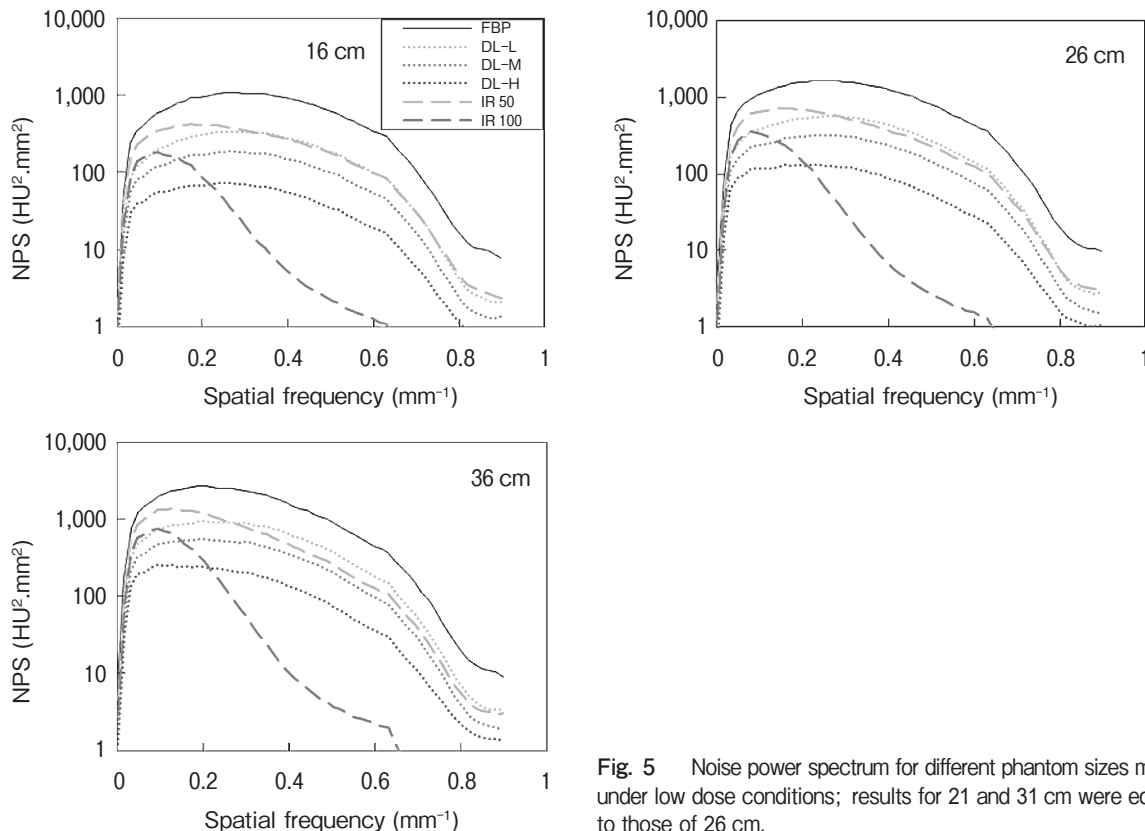


Fig. 5 Noise power spectrum for different phantom sizes measured under low dose conditions; results for 21 and 31 cm were equivalent to those of 26 cm.

Table 1 NPS_{peak} and NPS_{average} values with different phantom sizes for each dose condition

Phantom size (cm)	NPS peak (HU ² mm ²)											
	NI 13.7						NI 22.4					
	FBP	DL-L	DL-M	DL-H	IR 50	IR 100	FBP	DL-L	DL-M	DL-H	IR 50	IR 100
16	370	134	82	40	147	68	1088	340	188	73	428	182
21	563	209	127	64	221	105	1573	503	283	116	662	282
26	557	212	132	67	224	104	1642	561	316	132	707	359
31	520	198	126	65	219	104	1732	613	352	150	795	407
36	750	282	178	91	319	148	2646	926	543	258	1375	755

Phantom size (cm)	NPS average spatial frequency (mm ⁻¹)											
	NI 13.7						NI 22.4					
	FBP	DL-L	DL-M	DL-H	IR 50	IR 100	FBP	DL-L	DL-M	DL-H	IR 50	IR 100
16	0.33	0.32	0.31	0.30	0.29	0.14	0.33	0.32	0.32	0.30	0.29	0.14
21	0.33	0.31	0.30	0.29	0.29	0.14	0.32	0.32	0.31	0.30	0.28	0.14
26	0.32	0.31	0.30	0.28	0.28	0.14	0.31	0.31	0.30	0.28	0.27	0.13
31	0.31	0.30	0.29	0.27	0.27	0.14	0.30	0.30	0.29	0.28	0.26	0.13
36	0.31	0.30	0.29	0.27	0.27	0.14	0.28	0.29	0.28	0.26	0.24	0.13

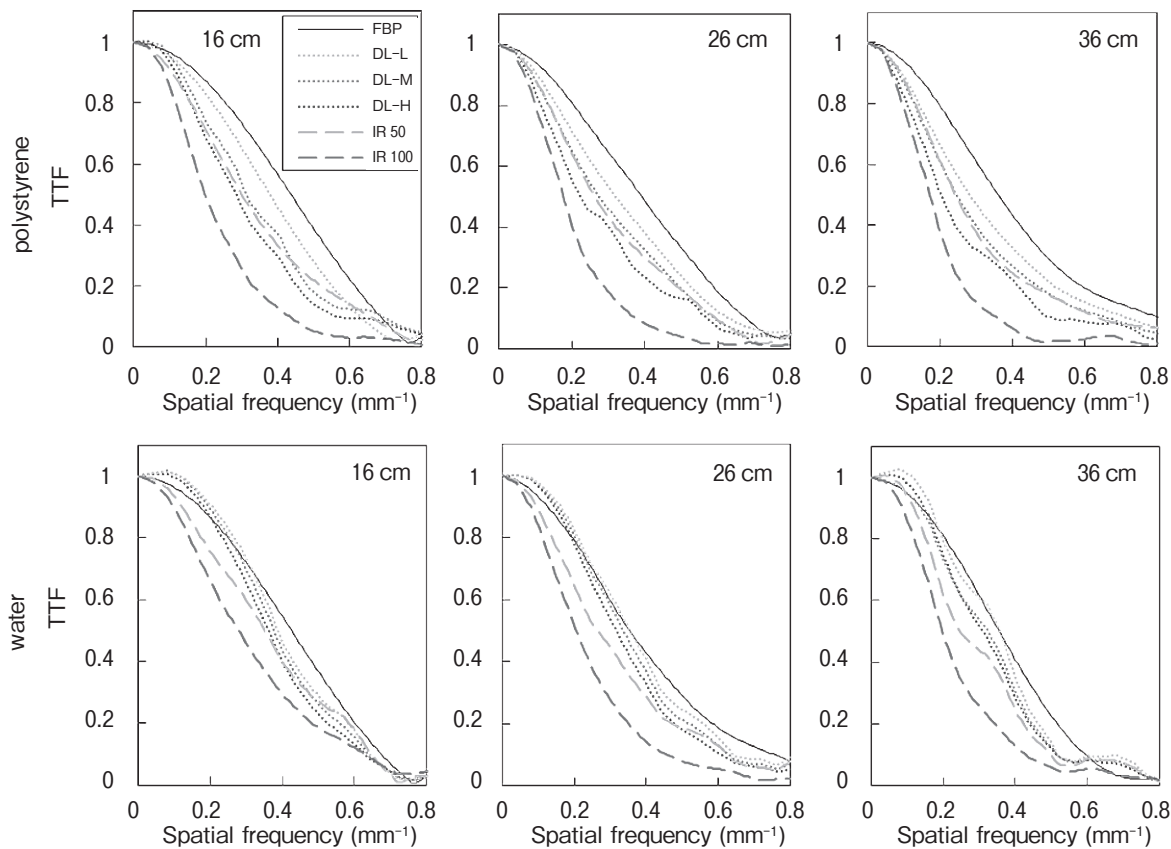


Fig. 6 Task-based modulation transfer functions of polyethylene and water with different phantom sizes at low dose conditions; this trend was maintained regardless of radiation dose.

with FBP, both DLIR and ASIR-V decreased the TTF50% as the reconstruction strength increased. The case with a TTF50%-decrease rate with respect to FBP also exhibited the largest decrease rate of DL-H, with a maximum of 44.4% (36 cm in diameter, NI=22.4). In hybrid-IR, the rate of decrease in IR100 was the highest, reaching a maximum of 59.1% (21 cm in diameter, NI=22.4). In TTF water, FBP and DLIR were similar, and DLIR exhibited favorable results in the low-spatial-frequency region. The results obtained from water were similar to those of polystyrene; and compared with FBP, both DLIR and ASIR-V showed decreased TTF50% as the reconstruction strength increased. In DLIR, the decrease rate of TTF50% with respect to FBP was the largest for DL-H, with a maximum of 16.3%

(16 cm in diameter, NI=22.4). In hybrid-IR, the rate of decrease was the highest for IR100, with a maximum of 44.4% (36 cm in diameter, NI=22.4).

Fig. 7 shows the d' (low-dose) results for each phantom size, and Table 3 shows the results of the increase in d' % at each dose level and phantom size. The trends in the results were similar regardless of the variation in dose, and FBP was the least detectable under all conditions. Compared with IR50, the detectabilities of DL-M and DL-H was high, while the detectability of DL-H were high under all conditions. The smaller the phantom size, the higher the detectability. However, there was no difference in d' by phantom size for the 21, 26, and 31 cm diameters, where the NPS_{peak} variations were small.

Table 2 Target transfer function at 50% for polystyrene and water inserts at different dose conditions and for each phantom size

		TTF 50% (mm ⁻¹)											
		NI:13						NI:22.4					
Phantom size (diameter)		FBP	DL-L	DL-M	DL-H	IR 50	IR 100	FBP	DL-L	DL-M	DL-H	IR 50	IR 100
Polystyrene (* Δ 50HU)	16	0.38	0.32	0.31	0.29	0.29	0.22	0.44	0.38	0.31	0.28	0.3	0.2
	21	0.41	0.32	0.3	0.28	0.28	0.21	0.44	0.34	0.3	0.25	0.28	0.18
	26	0.37	0.31	0.29	0.26	0.28	0.19	0.39	0.32	0.28	0.22	0.26	0.18
	31	0.35	0.3	0.28	0.26	0.26	0.19	0.31	0.27	0.24	0.22	0.23	0.18
	36	0.37	0.31	0.29	0.27	0.27	0.21	0.36	0.28	0.25	0.2	0.24	0.17
Water (* Δ 90HU)	16	0.38	0.39	0.38	0.38	0.37	0.35	0.43	0.39	0.38	0.36	0.35	0.28
	21	0.38	0.38	0.37	0.36	0.33	0.29	0.37	0.36	0.35	0.33	0.3	0.22
	26	0.36	0.36	0.35	0.34	0.31	0.27	0.35	0.35	0.34	0.32	0.27	0.21
	31	0.34	0.34	0.33	0.33	0.29	0.26	0.37	0.33	0.32	0.31	0.28	0.21
	36	0.35	0.35	0.35	0.34	0.3	0.26	0.36	0.36	0.32	0.31	0.24	0.2

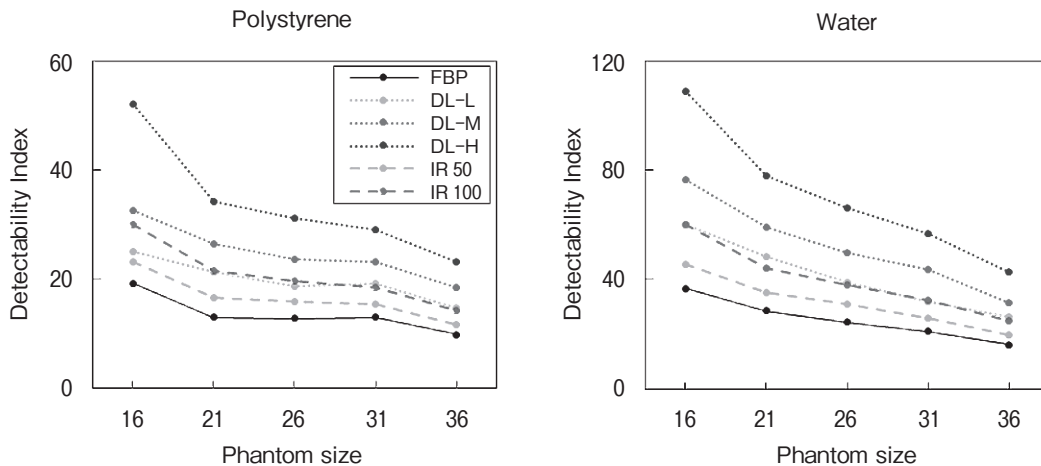


Fig. 7 Detectability index (d') based on phantom size for different contrasts (polystyrene and water) at low dose conditions.

Table 3 Rate of increase in detectability index (d' %) with each phantom size for different contrasts (polystyrene and water) at each dose condition

Phantom size (diameter)		Rate of increase in d' (%)									
		NI:13					NI:22.4				
		DL-L	DL-M	DL-H	IR 50	IR 100	DL-L	DL-M	DL-H	IR 50	IR 100
Polystyrene (* Δ 50HU)	16	40.0	76.8	124.2	20.5	48.7	30.2	69.6	170.2	20.2	55.7
	21	43.9	73.1	125.1	23.7	59.7	64.4	103.3	163.0	27.5	66.1
	26	38.0	60.2	103.3	17.2	44.7	47.4	84.9	143.8	23.8	54.7
	31	28.2	56.5	90.8	14.5	36.6	48.7	78.2	123.4	18.7	42.5
	36	21.9	62.5	100.0	20.6	45.9	50.2	89.5	136.7	18.7	44.7
Water (* Δ 90HU)	16	45.5	86.7	139.6	24.3	58.1	63.6	109.1	196.2	23.6	64.2
	21	41.7	72.4	125.6	21.1	57.8	69.3	108.0	172.5	23.6	54.4
	26	45.5	70.9	114.4	22.8	55.3	59.6	103.4	171.1	26.5	55.3
	31	32.3	65.8	103.1	15.7	41.5	50.5	106.3	169.0	22.8	52.8
	36	41.9	61.4	98.8	18.5	48.9	63.2	94.4	166.2	23.6	53.3

Regarding the images used for the observer study, Table 4 shows the results of the mean Hounsfield Units (HUs) and SDs measured by placing ROIs in the cervical muscles and abdominal liver of the five cases reconstructed with FBP. The mean SD values of the images were 21.8 and 38.0, respectively. Table 5 shows the average psychological measures and yardsticks of the five observers for the five cases. The calculated yardstick revealed that DL-H had the highest average psychological measures in all cases, regardless of the body part. In most cases, DL-H had the highest average psychological measure, followed by IR50, IR100, and FBP in that order; however, there was no significant difference between IR and FBP. For IR50 and IR100, the average psychological measures for IR50 were higher in the neck; however, the average psychological measures for IR100 were higher in the abdomen. The yardstick value varied slightly among cases, although it was comparable across body parts.

Discussion

The image quality of DLIR was evaluated using TCM with a Mercury 4.0 Phantom. The NPS_{peak} varied when the phantom size was 16 and 36 cm relative to the set NI. The NPS_{peak} values under each condition also varied accordingly, possibly because of the bowtie filter of the CT system [32]. The bowtie filter equalizes the in-plane dose distribution of the human body. In general, the size of the bowtie filter varies according to the size of the

Table 4 HU and SD of cervical muscle and abdominal liver in five cases reconstructed with FBP

		HU	SD
Neck	Pt. 01	49.8	16.1
	Pt. 02	34.3	20.9
	Pt. 03	33.4	23.4
	Pt. 04	38.6	23.4
	Pt. 05	40.2	25.4
	Ave.	39.3	21.8
Abdomen	Pt. 01	31.9	30.7
	Pt. 02	48.7	40.9
	Pt. 03	64.1	37.1
	Pt. 04	57.2	42.0
	Pt. 05	58.7	39.1
	Ave.	52.1	38.0

subject, but it cannot be changed during the scan. Therefore, we believe that this affected the control by TCM and caused the NPS_{peak} to vary.

According to the $NPS_{average}$ frequency and shape of the NPS, DLIR was able to maintain the shape and reduce the amount of image noise without any shift in the $NPS_{average}$ frequency compared to FBP. The ideal noise reduction, where the noise characteristics are improved by increasing the dose, occurs when the shape of the NPS is maintained, and the amount of shift is reduced, so that the noise reduction method of the DLIR algorithm can be considered to reduce noise without changing the details of the original image.

Table 5 Average psychological measures (95% confidence interval), and yardsticks for five observers obtained from a paired comparison observer study using five cases and two parts with different image reconstructions (FBP, DL-H, IR50, and IR100)

		Reconstruction method				Yardstick
		FBP [95% CI]	DL-H [95% CI]	IR 50 [95% CI]	IR 100 [95% CI]	
Neck	Pt. 01	-0.07 [-0.1148, -0.0252]	0.12 [0.0752, 0.1648]	-0.01 [-0.0548, 0.0348]	-0.04 [-0.0848, 0.0048]	0.0448
	Pt. 02	-0.06 [-0.1042, -0.0158]	0.09 [0.0458, 0.1342]	0.01 [-0.0342, 0.0542]	-0.04 [-0.0842, 0.0042]	0.0442
	Pt. 03	-0.05 [-0.0564, -0.0036]	0.11 [0.0636, 0.1564]	0.01 [-0.0364, 0.0564]	-0.06 [-0.1064, -0.0136]	0.0464
	Pt. 04	-0.08 [-0.1178, -0.0422]	0.13 [0.0922, 0.1678]	-0.02 [-0.0578, 0.0178]	-0.03 [-0.0678, 0.0078]	0.0378
	Pt. 05	-0.08 [-0.1291, -0.0309]	0.10 [0.0509, 0.1491]	0.01 [-0.0391, 0.0591]	-0.03 [-0.0791, 0.0191]	0.0491
	Ave.	-0.068	0.11	0.00	-0.04	
Abdomen	Pt. 01	-0.07 [-0.1110, -0.0290]	0.15 [0.1090, 0.1910]	-0.04 [-0.0810, 0.0010]	-0.04 [-0.0810, 0.0010]	0.0410
	Pt. 02	-0.09 [-0.1930, -0.0410]	0.14 [0.0910, 0.1890]	0.00 [-0.0490, 0.0490]	-0.05 [-0.0990, -0.0010]	0.0490
	Pt. 03	-0.05 [-0.0961, -0.0039]	0.15 [0.1039, 0.1961]	-0.05 [-0.0961, -0.0039]	-0.04 [-0.0861, -0.0139]	0.0461
	Pt. 04	-0.11 [-0.1439, -0.0761]	0.18 [0.1461, 0.2139]	-0.04 [-0.0739, -0.0061]	-0.04 [-0.0739, -0.0061]	0.0339
	Pt. 05	-0.10 [-0.1505, -0.0495]	0.16 [0.1095, 0.2105]	-0.04 [-0.0905, 0.0105]	0.02 [-0.0305, 0.0705]	0.0505
	Ave.	-0.084	0.156	-0.034	-0.03	

Conversely, the $NPS_{average}$ frequency of IR50 did not exhibit a large variation, while it did, similar to IR100, exhibit a characteristic shape as if it had been smoothed in the high-frequency region. The effect of smoothing on IR100 was strong, and the $NPS_{average}$ frequency also shifted significantly to the low-frequency side. The IR100 image also shows the influenced of high-frequency components on image quality (Fig. 8), which is consistent with previous research [5]. Regarding the TTF, the focus size was automatically changed by the NI setting, and the effect on TTF was expected, although very little effect was observed [33]. Previous studies on the TTF of DLIR have reported that the resolution characteristics are improved compared to the FBP, which was not necessarily the case in this experiment [8, 10]. Hara *et al.* [34] reported the effect of the ROI position on the measured MTF value. The effect of the selected phantom on the measurement results cannot be ignored. However, it is considered that DLIR easily recognizes high contrast, such as 300 HU, as reported in a previous study; however, at a low contrast of approximately 50 HU, as in this study, the resolution characteristics were lower than those of FBP.

Regarding d' , FBP was less detectable at all doses and phantom sizes. In hybrid-IR and DLIR, the detectability of the image reconstructed by DLIR was higher than that of IR50 under all conditions. DL-H exhibited the highest detectability, with an improvement in reconstruction strength. Regarding TTF, DLIR did not always show high-resolution characteristics due to con-

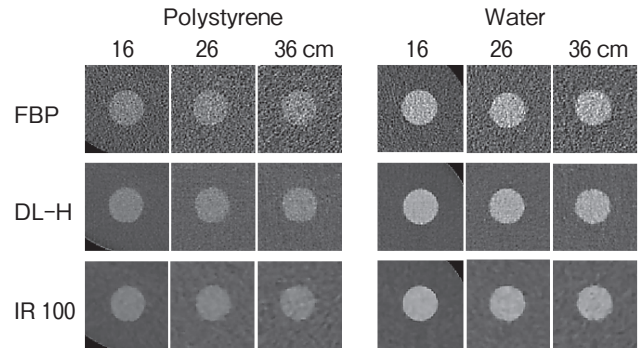


Fig. 8 Images of each contrast insert with phantom size at low dose conditions.

trast; however, NPS exhibited high noise characteristics (low image noise), and the simulated tumor revealed high d' when reconstructed by DLIR. The simulated tumor size was 8 mm, which translates to approximately 0.06 mm^{-1} in terms of spatial frequency, and the difference in NPS of each reconstruction method at 0.06 mm^{-1} was significant, which we believe had a significant impact on the results of d' . Moreover, Urikura *et al.* [20] reported that the effect of the resolution characteristics was small for such low-contrast visibility, and considered that the high noise reduction effect of DLIR was highly useful in clinical practice. For phantom sizes of 16 and 36 cm, d' varies greatly, indicating the effect of noise characteristics.

The results of the observer study using clinical cases showed different characteristics for IR and DLIR. As in

the physical evaluation, the average psychological measures for images reconstructed with DLIR were the highest, with a significant difference regardless of the subject size. The difference in average psychological measures between FBP and DL-H and the difference between FBP and IR50 in the neck was 0.178 and 0.068, respectively. The differences in average psychological measures between FBP and DL-H and between FBP and IR50 in the abdomen were 0.24 and 0.05, respectively. As previously reported [5], the image quality of IR is known to be affected by the noise level of the images, and we believe that the lower evaluation of IR compared to FBP is due to the higher noise level of the images in the abdomen compared to the neck. However, DLIR exhibited high average psychological measures in the observer study, as well as in the physical evaluation. Based on our results, we believe that noise reduction without bias in the frequency range reproduced the natural image quality in clinical images. Furthermore, a clinical evaluation by Jensen *et al* [12] reported that DLIR reduces artifacts as well as noise. Similarly in this study, streak artifacts caused by insufficient dose in the abdominal images were improved by DLIR. This may explain the high average psychological measures for DLIR.

Moreover, there was no significant difference between the neck and abdomen for Yardstick. This indicates that the observer variation was small regardless of the body part, and that a uniform evaluation was obtained by the observer. Under conditions such as a torso examination, which is assumed to be a clinical situation, the subject size may vary greatly, and in such a case, the hardware of the device may be affected by the bowtie filter and output limitations. In such cases, the high noise reduction performance of DLIR compared to FBP and IR can provide stable image quality that is less affected by subject size and noise level.

The present study had several limitations. First, it relied heavily on the capabilities of commercially available equipment. The tube voltage, which represents the scan condition, was only 120 kVp, and all scans were performed helically. Second, the analysis using different phantom sizes was performed based on adult size; the smaller phantom sizes did not represent children. The phantom is also circular and not an ellipse, which better imitates the conditions of a real person. The arrangement of the measured ROIs was the same regardless of the phantom size when measuring the NPS

and TTF, which was necessary to maintain the accuracy of the experiment, although previous studies have described the effect of the measurement position. Finally, the cases used in the observer experiment were obtained under specific clinical conditions, and observations were made only in the transverse section.

In conclusion, we evaluated the overall image quality assessment of DLIR when TCM was used, targeting low-to-medium contrasts and varying phantom sizes and NIs. The results of noise control by TCM revealed that DLIR has a high detectability owing to its high noise reduction capability, regardless of the phantom size. The same trend was observed in the observer study, with DLIR showing a higher average psychological measure than FBP and IR. Depending on the patient size, the hardware of the CT system may be restricted for clinical use; however, the DLIR algorithm can provide stable images independent of the object.

Acknowledgments. We would like to thank Editage (www.editage.com), KN International for English language editing.

References

1. OECD: Health at a Glance 2015: OECD Indicators; OECD Publishing, Paris (2015).
2. Brenner DJ and Hall EJ: Computed tomography—an increasing source of radiation exposure. *N Engl J Med* (2007) 357: 2277–2284.
3. Geyer LL, Schoepf UJ, Meinel FG, Nance Jr JW, Bastarrika G, Leipsic JA, Paul NS, Rengo M, Laghi A and De Cecco CN: State of the Art: Iterative CT Reconstruction Techniques. *Radiology* (2015) 276: 339–357.
4. Löve A, Olsson M and Siemund R: Six iterative reconstruction algorithms in brain CT: a phantom study on image quality at different radiation dose levels. *Br J Radiol* (2013) 86 (1031): 20130388.
5. McCollough CH, Yu L, Koffler JM, Leng S, Zhang Y, Li Z and Carter RE: Degradation of CT Low-Contrast Spatial Resolution Due to the Use of Iterative Reconstruction and Reduced Dose Levels. *Radiology* (2015) 276: 499–506.
6. Mileto A, Guimaraes LS, McCollough CH, Fletcher JG and Yu L: State of the Art in Abdominal CT: The Limits of Iterative Reconstruction Algorithms. *Radiology* (2019) 293: 491–503.
7. Urikura A, Hara T, Ichikawa K, Nishimaru E, Hoshino T, Yoshida T and Nakaya Y: Objective assessment of low-contrast computed tomography images with iterative reconstruction. *Phys Med* (2016) 32: 992–998.
8. Greffier J, Hamard A, Pereira F, Barrau C, Pasquier H, Beregi JP and Frandon J: Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. *Eur Radiol* (2020) 30: 3951–3959.
9. Hsieh J, Liu E, Nett B, Tang J, Thibault JB and Sahney S: A new era of image reconstruction: TrueFidelity™. Technical white paper on deep learning image reconstruction. GE Healthcare (2019)

10. Racine D, Becce F, Viry A, Monnin P, Thomsen B, Verdun FR and Rotzinger DC: 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* (2020) 76: 28–37.
11. Benz DC, Benetos G, Rampidis G, Felten EV, Bakula A, Sustar A, Kudura K, Messerli M, Fuchs TA, Gebhard C, Pazhenkottil AP, Kaufmann AP and Buechel RR: Validation of deep-learning image reconstruction for coronary computed tomography angiography: Impact on noise, image quality and diagnostic accuracy. *J Cardiovasc Comput* (2020) 14: 444–451.
12. Jensen CT, Liu X, Tamm EP, Chandler AG, Sun J, Morani AC, Javadi S and Wagner-Bartak NA: Image Quality Assessment of Abdominal CT by Use of New Deep Learning Image Reconstruction: Initial Experience. *Am J Roentgenol* (2020) 215: 50–57.
13. Higaki T, Nakamura Y, Zhou J, Yu Z, Nemoto T, Tatsugami F and Awai K: Deep Learning Reconstruction at CT: Phantom Study of the Image Characteristics. *Acad Radiol* (2020) 27: 82–87.
14. Solomon JB, Li X and Samei E: Relating noise to image quality indicators in CT examinations with tube current modulation. *Am J Roentgenol* (2013) 200: 592–600.
15. Solomon J, Lyu P, Marin D and Samei E: Noise and spatial resolution properties of a commercially available deep learning-based CT reconstruction algorithm. *Med Phys* (2020) 47: 3961–3971.
16. Ria F, Solomon JB, Wilson JM and Samei E: Technical Note: Validation of TG 233 phantom methodology to characterize noise and dose in patient CT data. *Med Phys* (2020) 47: 1633–1639.
17. Takata T, Ichikawa K, Mitsui W, Hayashi H, Minehiro K, Sakuta K, Nunome H, Matsubara K, Kawashima H, Matsuura Y and Gabata T: Object shape dependency of in-plane resolution for iterative reconstruction of computed tomography. *Phys Medica* (2017) 33: 146–151.
18. Furlan A, Marin D, Vanzulli A, Patera GP, Ronzoni A, Midiri M, Bazzocchi M, Lagalla R and Brancatelli G: Hepatocellular carcinoma in cirrhotic patients at multidetector CT: hepatic venous phase versus delayed phase for the detection of tumour washout. *Brit J Radiol* (2011) 84: 403–412.
19. Cham MD, Yankelevitz DF, Shaham D, Shah AA, Sherman L, Lewis A, Rademaker J, Pearson G, Choi J, Wolff W, Prabhu PM, Galanski M, Clark RA, Sostman HD and Henschke CI: Deep venous thrombosis: detection by using indirect CT venography. The Pulmonary Angiography-Indirect CT Venography Cooperative Group. *Radiology* (2000) 226: 744–751.
20. Urikura A, Ichikawa K, Hara T, Nishimaru E and Nakaya Y: Spatial resolution measurement for iterative reconstruction by use of image-averaging techniques in computed tomography. *Radiol Phys Technol* (2014) 7: 358–366.
21. Ria F, Solomon JB, Wilson JM and Samei E: Technical Note: Validation of TG 233 phantom methodology to characterize noise and dose in patient CT data. *Med Phys* (2020) 47 (4): 1633–1639.
22. Samei E, Bakalyar D, Boedeker KL, Brady S, Fan J, Leng S, Myers K, Popescu LM, Giraldo JCR, Ranallo F, Solomon J, Vaishnav J and Wang J: Performance evaluation of computed tomography systems: Summary of AAPM Task Group 233. *Med Phys* (2019) 46: e735–e756.
23. Gies M, Kalender WA, Wolf H and Suess C: Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies. *Med Phys* (1999) 26 (11): 2235–2247.
24. IAEA: IAEA Safety Series No.1154, International Basic Safety Standards for Protection of Radiation Sources. IAEA, Vienna (1994).
25. Richard S, Husarik DB, Yadava G, Murphy SN and Samei E: Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. *Med Phys* (2012) 39: 4115–4122.
26. Yu L, Vrieze TJ, Leng S, Fletcher JG and McCollough CH: Technical Note: Measuring contrast- and noise-dependent spatial resolution of an iterative reconstruction method in CT using ensemble averaging. *Med Phys* (2015) 42: 2261–2267.
27. Solomon J and Samei E: Correlation between human detection accuracy and observer model-based image quality metrics in computed tomography. *J Med Imaging* (2016) 3: 035506.
28. Chen YL, Ko CJ, Chien SY, Sheng L, Chen CM, Chi C and Lai H: Tumor size as a prognostic factor in resected small hepatocellular carcinoma: a controversy revisited. *J Gastroen Hepatol* (2011) 26: 851–857.
29. Shiraishi J, Okazaki Y and Goto M: Image Evaluation with Paired Comparison Method Using Automatic Analysis Software: Comparison of CT Images with Simulated Levels of Exposure Dose. *Nihon Hoshasen Gijutsu Gakkai Zasshi* (2019) 75(1): 24–31.
30. H Scheffé: An analysis of variance for paired comparisons. *J Am Stat Assoc* (1952) 47: 381–400.
31. Metz CE, Herman BA and Shen JH: Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med* (1998) 17: 1033–1053.
32. Gomez-Cardona D, Cruz-Bastida JP, Li K, Budde A, Hsieh J and Chen G: Impact of bowtie filter and object position on the two-dimensional noise power spectrum of a clinical MDCT system. *Med Phys* (2016) 43: 4495–4506.
33. Grimes J, Duan X, Yu L, Halaweish AF, Haag N, Leng S and McCollough C: The influence of focal spot blooming on high-contrast spatial resolution in CT imaging. *Med Phys* (2015) 42: 6011–6020.
34. Hara T, Ichikawa K, Sanada S and Ida Y: Image quality dependence on in-plane positions and directions for MDCT images. *Eur J Radiol* (2010) 75: 114–121.