



# Dose reduction potential of vendor-agnostic deep learning model in comparison with deep learning–based image reconstruction algorithm on CT: a phantom study

Hyunsu Choi<sup>1</sup> · Won Chang<sup>1</sup> · Jong Hyo Kim<sup>2,3</sup> · Chulkyun Ahn<sup>3</sup> · Heejin Lee<sup>4</sup> · Hae Young Kim<sup>1</sup> · Jungheum Cho<sup>1</sup> · Yoon Jin Lee<sup>1</sup> · Young Hoon Kim<sup>1</sup>

Received: 29 March 2021 / Revised: 11 June 2021 / Accepted: 2 July 2021 / Published online: 14 August 2021  
© European Society of Radiology 2021

## Abstract

**Objectives** To compare the dose reduction potential (DRP) of a vendor-agnostic deep learning model (DLM, ClariCT.AI) with that of a vendor-specific deep learning–based image reconstruction algorithm (DLR, TrueFidelity™).

**Methods** Computed tomography (CT) images of a multi-sized image quality phantom (Mercury v4.0) were acquired under six radiation dose levels (0.48/0.97/1.93/3.87/7.74/15.47 mGy) and were reconstructed using filtered back projection (FBP) and three strength levels of the DLR (low/medium/high). The FBP images were denoised using the DLM. For all DLM and DLR images, the detectability index ( $d'$ ) (a task-based detection performance metric) was obtained, under various combinations of three target sizes (10/5/1 mm), five inlets (CT value difference with the background; –895/50/90/335/1000 HU), five phantom diameters (36/31/26/21/16 cm), and six radiation dose levels. Dose reduction potential (DRP) measures the dose reduction made by using DLM or DLR, while yielding  $d'$  equivalent to that of FBP at full dose.

**Results** The DRPs of the DLM, DLR-low, DLR-medium, and DLR-high were 86% (81–88%), 60% (46–67%), 76% (60–81%), and 87% (78–92%), respectively. For 10-mm targets, the DRP of the DLM (87%) was higher than that of all DLR algorithms (58–86%). However, for smaller targets (5 mm/1 mm), the DRPs of the DLR-high (89/88%) were greater than those of the DLM (87/84%).

**Conclusion** The dose reduction potential of the vendor-agnostic DLM was shown to be comparable to that of the vendor-specific DLR at high strength and superior to those of the DLRs at medium and low strengths.

## Key Points

- DRP of the vendor-agnostic model was comparable to that of high-strength vendor-specific model and superior to those of medium- and low-strength models.
- Under various radiation dose levels, the deep learning model shows higher detectability indexes compared to FBP.

**Keywords** Deep learning · Tomography, x-ray computed · Phantoms, imaging · Radiation dosage · Artificial intelligence

## Abbreviations

CNR Contrast-to-noise ratio

CT Computer tomography

CTDI CT dose index

✉ Won Chang  
changwon1981@gmail.com

<sup>1</sup> Department of Radiology, Seoul National University Bundang Hospital, 82, Gumi-ro-173-beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620, Republic of Korea

<sup>2</sup> Department of Radiology, Seoul National University College of Medicine, and Institute of Radiation Medicine, Seoul National University Medical Research Center, 101 Daehak-ro, Jongno-gu, Seoul, Republic of Korea

<sup>3</sup> Department of Transdisciplinary Studies, Program in Biomedical Radiation Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

<sup>4</sup> Department of Applied bioengineering, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

d'	Detectability index
DLM	Deep learning model
DLR	Deep learning-based image reconstruction
DRP	Dose reduction potential
FBP	Filtered back projection
GAN	General adversarial network
IR	Iterative reconstruction
NPS	Noise power spectrum
NPW	Non-pre whitening
TG	Task force
TTF	Task-based transfer function

## Introduction

Although computed tomography (CT) is a widely used imaging modality in clinical practice [1], it inherently involves the hazard of radiation exposure [2]. Iterative reconstruction (IR) has facilitated technical efforts to reduce the radiation dose without compromising diagnostic performance [3–5].

Recently, however, deep learning-based image reconstruction algorithms (DLRs) have emerged as an alternative technical approach. For example, a DLR introduced by GE Healthcare (TrueFidelity™) was reported to reduce the radiation dose by 36–50% compared with IR, while decreasing image noise and improving lesion detectability [6–10]. However, because it is a vendor-specific algorithm, it can only be mounted and operated on state-of-the-art GE Healthcare machines. The need for a dose reduction technique that is applicable to any existing CT scanner was strongly reinforced by the coronavirus disease 2019 pandemic, with the increased demand for chest CTs [11]. Therefore, the development and evaluation of vendor-agnostic deep learning-based algorithms that improve image quality with a lower radiation dose is a timely and necessary task.

A few vendor-agnostic techniques that can reduce image noise with non-inferior detectability compared to existing vendor-specific IR and DLR have been introduced [8, 12–14]. ClariCT.AI (ClariPI) is a vendor-agnostic deep learning model (DLM) that works in the image post-processing domain and does not need projection data. According to a recent study by Nam et al, for ultralow-dose CT, the DLM achieved better subjective image quality with fewer artifacts than did the DLR, although the vendor-specific DLR showed less noise and a higher signal-to-noise ratio than did the vendor-agnostic DLM [15]. However, this study assessed the subjective image quality under a single radiation dose level, and the diagnostic performance was not evaluated.

With the development of various techniques (e.g., IR), for CT, AAPM Task Group (TG) 233 suggested a guideline for the performance evaluation of CT images. This guideline includes a detectability index, which is a task-based detection performance metric that indicates the clinical performance in detecting lesions [16]. To the best of our knowledge, the dose

reduction potentials of vendor-agnostic DLM and DLR have not yet been compared using the detectability index.

Therefore, in this study, we compared the dose reduction potential (DRP) of the vendor-agnostic DLM and DLR using the detectability index.

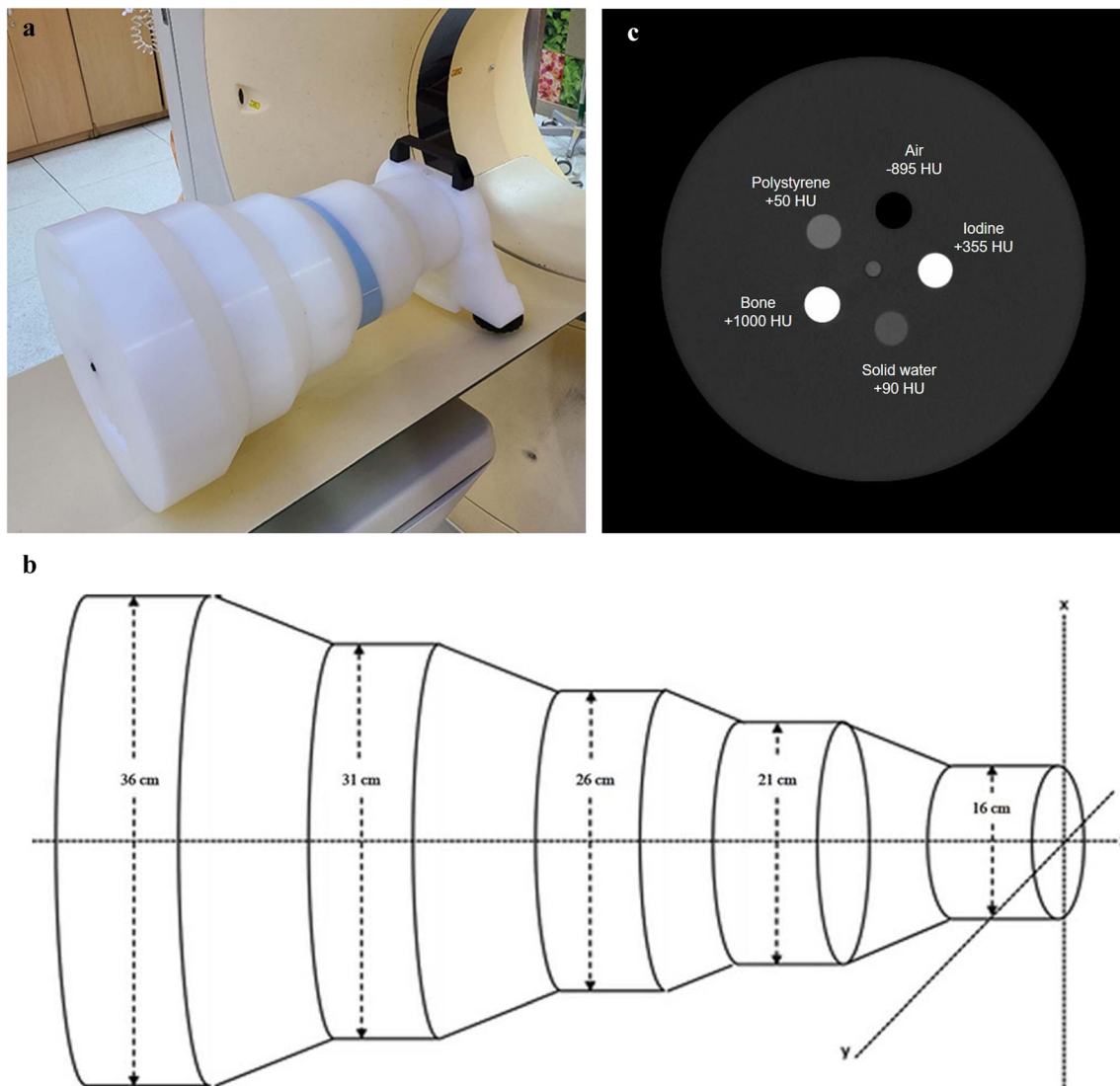
## Materials and methods

### Data acquisition and deep learning-based denoising algorithm

We used a multi-sized image quality phantom (Mercury v4.0 phantom, GAMMEX Sun Nuclear Corp.) suggested by AAPM TG 233 for assessing a CT system; it is a cylindrical polyethylene phantom with expected background CT values of −90 HU under 120 kVp. It consists of five sections with different diameters (36/31/26/21/16 cm) and a total length of 52 cm. Each section includes five inlets representing solid water, bone-mimicking material, polystyrene, 10-mg/ml iodine, and air (Fig. 1), each with expected CT difference with the background of 90, 1000, 50, 335, and −895 HU, respectively [16]. We measured the background CT value by drawing a concentric circular ROI around each inlet, excluding the area of the inlet (Supplementary Appendix Figure E1).

Images were acquired using Revolution CT (GE Healthcare) equipped with the TrueFidelity algorithm. Scanning was done under six radiation dose levels (volume CT dose index (CTDI<sub>vol</sub>): 15.47, 7.74, 3.87, 1.93, 0.97, and 0.48 mGy) with a tube voltage of 120 kV. We set the full-dose level in this study to 15.47 mGy by referencing the diagnostic reference level of abdominal pelvic CT, which is typically set at the 75th percentile of CTDI<sub>vol</sub> for patients with water equivalent diameters between 29 and 33 cm [17]. The applied CT parameters are listed in Table 1.

CT images of the Mercury phantom were reconstructed using the FBP and a DLR (TrueFidelity-L/M/H). The inference engine of TrueFidelity generates three selectable reconstruction strength levels (i.e., low, medium, high (L/M/H)) to control the degree of noise reduction. Reconstruction strength levels are qualitatively divided according to the noise reduction and are selectable based on the clinical application. A body kernel was used for reconstruction with FBP. Scanning was performed 10 times at each radiation dose level [16, 18]. To achieve a reliable estimate of the task-based transfer function (TTF), which represents the spatial resolution of the CT image, the contrast-to-noise ratio (CNR) should be at least 15. However, for low-contrast inlets with high image noise, the minimum CNR may not be achievable in a single slice. In that case, multiple slices should be averaged to lower the noise and increase the CNR [16, 18]. We followed the averaging method specified by AAPM TG 233 to achieve a total effective CNR greater than 15 [16, 18]. We obtained TTF to calculate a metric representing the CT image quality (i.e., detectability index), which will be described in detail later.



**Fig. 1** **a** Mercury phantom which was used in the study. **b** Schematic of Mercury phantom (v4.0) consisted of five sections of different diameters with tapered transitional sections. **c** Three-centimeter-sized five inlets were radially distributed equally from the center in each section of diameter

Deep learning–based denoising from the FBP images was performed using the DLM [19]. The DLM is an image-based CT denoising solution having a U-net-based CNN, trained by taking a noise-added CT image as input to produce an original CT image as output. To generate the training dataset, a synthetic sinogram-based low-dose CT simulation technique was applied to create a paired set of low-dose and standard-dose CT images from a set of given standard-dose CT images [20]. To solve the problem of overfitting and to train a model that minimizes bias, we trained the DLM using various combinations of clinical and phantom images from four major vendors. Various vendor-specific datasets were used for the training, so that the DLM could acquire vendor-agnostic denoising capability. The training dataset comprised more than 1 million clinical CT images encompassing 2,100 different combinations of scan and reconstruction conditions, including various kVp, mAs,

automatic exposure controls, slice thicknesses, contrast enhancements, and convolution kernels, with 24 scanner models from four different CT manufacturers (GE Healthcare, Siemens, Philips Healthcare, and Canon). In this study, we used the DLM version (ver.1.2.1), which was further trained using the images of standard QA phantoms such as AAPM, ACR, and Catphan considering phantom-based image quality assurance practices. The DLM performance has been previously evaluated in several clinical studies [12–15, 21].

### Task-based image quality assessment

We used a dedicated analysis software tool (imQuest, Duke University) to evaluate the detectability index ( $d'$ ) [22]. This index is a metric of the image quality that is specifically designed to represent the detectability of target objects; an

**Table 1** CT scan parameters for scanning Mercury phantom v4.0

Parameter	Data
Tube potential (kVp)	120
Automated tube current modulation	off
Rotation time (s)	0.7
Collimation (mm)	64 × 0.625
Section thickness (mm)	0.625
Pitch (mm)	0.994
Scan field of view (cm)	50
Display field of view (cm)	38.2
Scan option	Helical
No. of pixel	512 pixels × 512 pixels
CTDIvol (mGy)	0.48/0.97/1.93/3.87/7.74/15.47
Tube current (mA)	10/20/40/80/160/320
Focal spot size	small

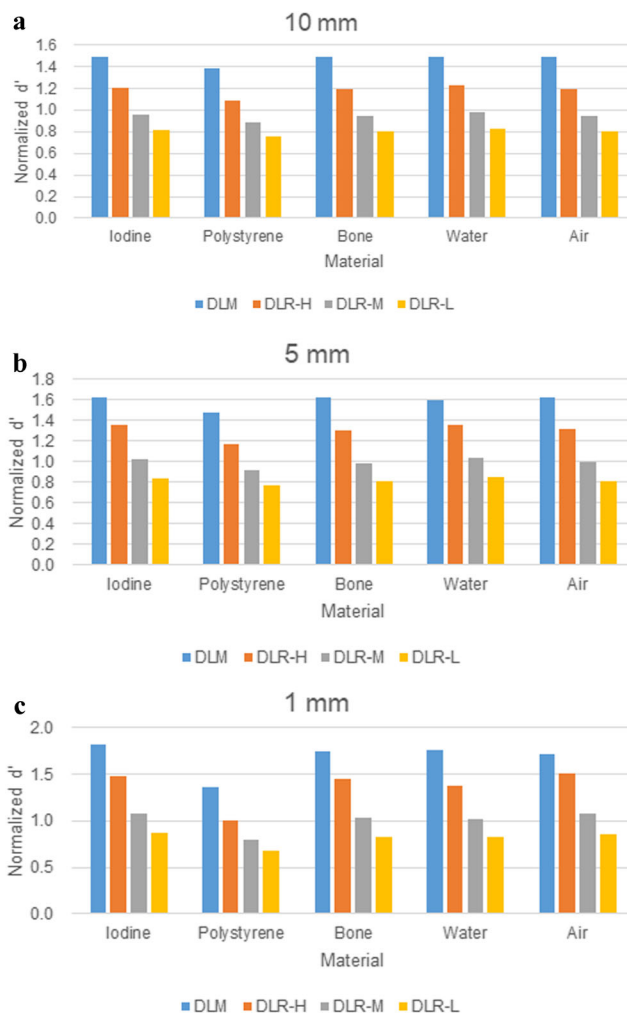
*DLM* deep learning-based model, *DLR* deep learning-based image reconstruction algorithm

increase in  $d'$  suggests that the target object is easier to detect, and the image quality is better [16]. The detectability index consists of three primary components: a task to be performed, an observer to perform the task, and the image to be evaluated. The task is to detect a circular signal defined by a task function ( $W_{\text{task}}$ ).  $W_{\text{task}}$  encodes the properties of the image (i.e., size, shape, contrast, and contrast profile), and these encoded profiles are used to calculate the detectability index. We used the target signals of designer profile with three sizes of circular signals (i.e., 1, 5, and 10 mm) for each of the five inlets. For the observer model, a non-pre-whitening (NPW) matched filter was adopted that correlates strongly with human performance in the low-contrast detection task [16].

To acquire the detectability index, the noise power spectrum (NPS) and TTF were first measured. The noise was measured in terms of the NPS to analyze the components of the Mercury phantom at each dose level. The TTF was measured as an indicator of the spatial resolution for each dose level and inlet. The spatial frequency at which the TTF reached 50% was indicated as  $\text{TTF}_{50\%}$  and was used to represent the system resolution under the given condition [16]. The change in  $\text{TTF}_{50\%}$  compared with the FBP was obtained with respect to the object and radiation dose. NPS and a detectability index were obtained according to the methods specified by AAPM task group 233.

The  $d'$  was obtained for DLM and DLR images using the methods introduced by AAPM TG 233 under various combinations of five inlets, five phantom diameters, and six radiation dose levels [16].

For convenience, the normalized  $d'$  value was used in consideration of the wide variation in  $d'$  across the five inlets. The normalization of  $d'$  at each dose level was achieved by dividing the value by that of the same inlet on the full dose (15.47 mGy) FBP image. For example, if the  $d'$  of detecting



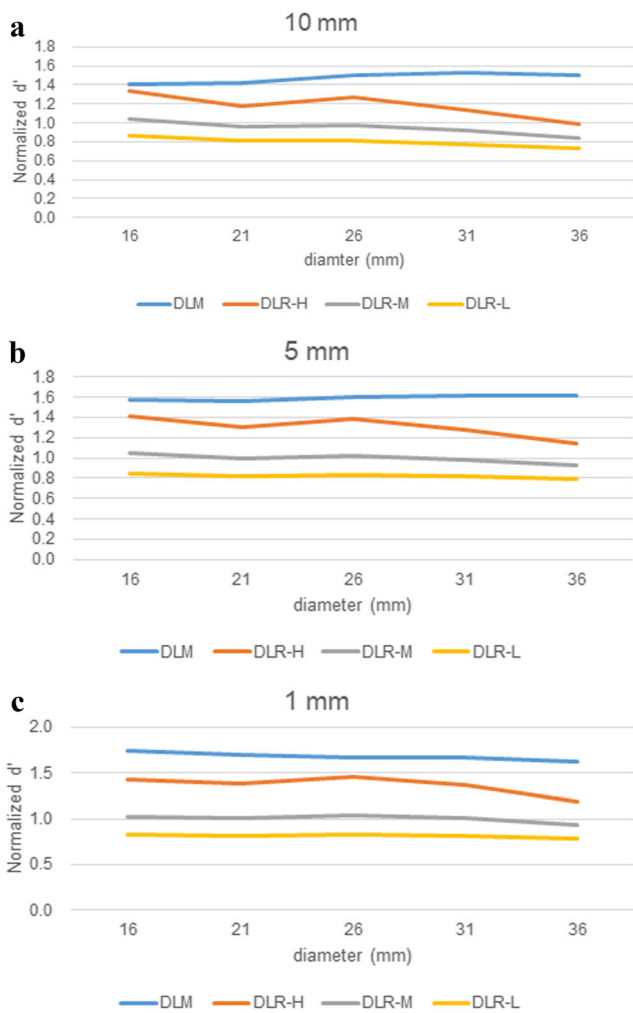
**Fig. 2** Normalized detectability index ( $d'$ ) of the DLM and three levels of the DLRs according to five inlets for 10-mm (a), 5-mm (b), and 1-mm (c) target objects. *DLM* deep learning model, *DLR-H* deep learning-based image reconstruction algorithm, high strength, *DLR-M* deep learning-based image reconstruction algorithm, medium strength, *DLR-L* deep learning-based image reconstruction algorithm, low strength

polystyrene target on the image acquired under 7.74 mGy and denoised with the DLM was 63, and if  $d'$  of that on the full-dose FBP image was 28, the normalized  $d'$  would be  $63/28 = 2.25$ . The formula for normalized  $d'$  was as follows:

$$\text{Normalized } d' = \frac{(\text{d' of the specific material at a specific radiation dose for the DLR or DLM algorithm})}{(\text{d' of the same material on full-dose FBP image})}$$

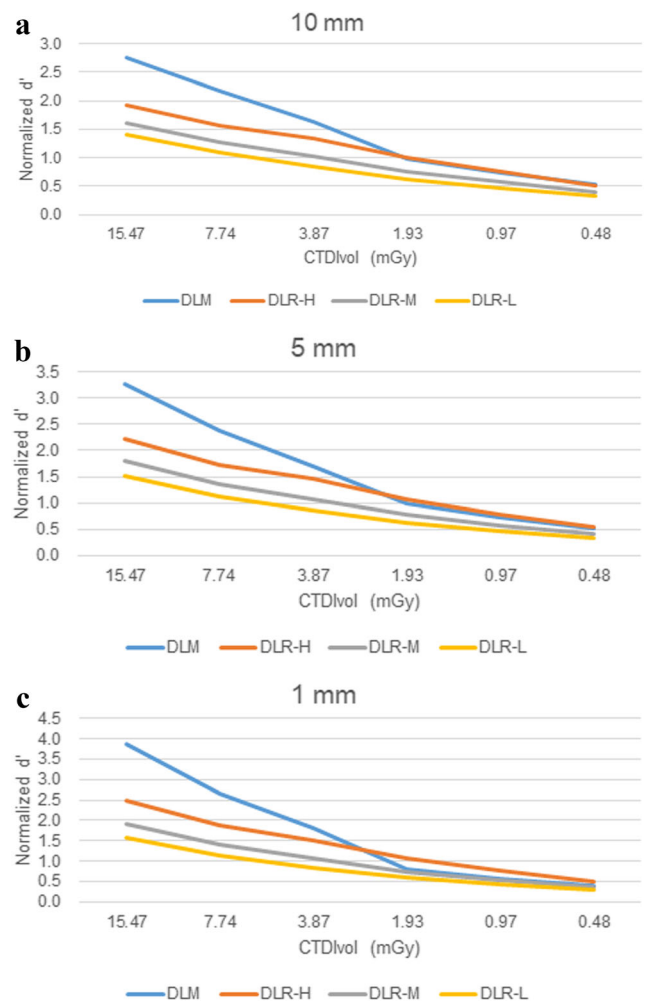
## Dose reduction potential

The performances of the DLM and DLR-L/M/H were compared by obtaining the DRP as the final comparison value.



**Fig. 3** Normalized detectability index ( $d'$ ) of the DLM and three levels of the DLRs according to phantom diameters for 10-mm (a), 5-mm (b), and 1-mm (c) target objects *DLM* deep learning model, *DLR-H* deep learning-based image reconstruction algorithm, high strength, *DLR-M* deep learning-based image reconstruction algorithm, medium strength, *DLR-L* deep learning-based image reconstruction algorithm, low strength

The DRP was defined as the amount of radiation dose that could be reduced by using either the DLM or DLR instead of the FBP. For example, the radiation dose at which the  $d'$  of polystyrene reaches 1 is 16 mGy for the FBP, whereas that for the DLM is 6 mGy. Thus, the DRP of the DLM is  $(16-6)$  mGy/ $16$  mGy  $\times 100$  (%) = 62.5 (%) [23]. For comparison, DRP values were averaged across the phantom size and inlets for each reconstruction algorithm (i.e., DLM, DLR-L/M/H). For example, the radiation dose at which the  $d'$  of 10-mm polystyrene target on DLM images became equal to that on FBP images acquired under the full dose (15.47 mGy) was 1.76, 1.93, 2.23, 2.29, and 2.67 mGy respectively, in 16-, 21-, 26-, 31-, and 36-cm phantoms. We obtained the average of those values, which was 2.18 mGy. We repeated the same process for the other four inlets, and the averaged values for the iodine, bone, water, and air were 1.89 mGy, 1.86 mGy, 1.94 mGy, and 1.87 mGy, respectively. Finally, the average of the



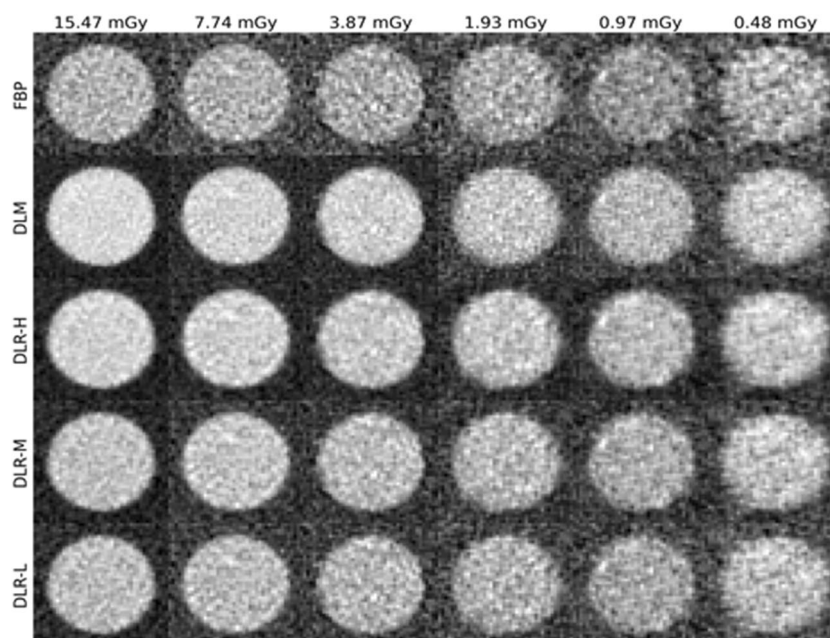
**Fig. 4** Normalized detectability index of the DLM and three levels of the DLRs according to radiation dose levels for 10-mm (a), 5-mm (b), and 1-mm (c) target objects *DLM* deep learning model, *DLR-H* deep learning-based image reconstruction algorithm, high strength, *DLR-M* deep learning-based image reconstruction algorithm, medium strength, *DLR-L* deep learning-based image reconstruction algorithm, low strength

values from the five inlets (1.95 mGy) was used for the calculation of the DRP. The DRP of the DLM in detecting 10 mm targets in this case was  $(15.47-1.95/15.47) \times 100 = 87\%$ .

## Results

The detectability index was analyzed with respect to the inlet, phantom diameter, and radiation dose. For all inlets and phantom diameters,  $d'$  of the DLM was higher than those of all DLR levels (Figs. 2 and 3). The difference in  $d'$  between the DLM and the DLR-H was greater when the diameter of the phantom was larger. At higher dose levels (3.87/7.74/15.47 mGy),  $d'$  of the DLM was higher than that of the DLR-H. For radiation doses below 1.93 mGy,  $d'$  of the DLM was lower than that of the DLR-H but was comparable to that of the DLR-M (Fig. 4). Figure 5 illustrates the image of polystyrene

**Fig. 5** A  $3 \times 3$  cm<sup>2</sup> region of interest centered on the polystyrene on the FBP, DLM, and DLR-H, DLR-M, DLR-L according to the radiation dose on the phantom diameter of 31 cm. *FBP* filtered back projection, *DLM* deep learning model, *DLR-H* deep learning-based image reconstruction algorithm, high strength, *DLR-M* deep learning-based image reconstruction algorithm, medium strength, *DLR-L* deep learning-based image reconstruction algorithm, low strength



object obtained by using different algorithms at each radiation dose level. For each algorithm, the CT value differences between the background and each inlet according to radiation dose and phantom diameter were presented in the supplementary appendix ([Supplementary Appendix Table 1–3](#)).

The average DRPs of the DLM, DLR-L, DLR-M, and DLR-H were 86% (81–88%), 60% (46–67%), 76% (60–81%), and 87% (78–92%), respectively. For large target objects (10 mm), the DRP of the DLM was slightly higher than those of all DLR levels. However, for smaller targets (1 mm and 5 mm), the DRP of the DLR-H (89 and 88%) was greater than that of the DLM (87 and 84%) (Table 2). The DLM showed a relatively narrow range of DRP with the range of 81–88%, in comparison with the DLR-H that showed the range of 78–92%. The CT values of the five inlets and the background, as well as the NPS and TTF results obtained to calculate  $d'$  are presented in the supplementary appendix ([Supplementary Appendix Fig. E2–7](#) and Tables 1–5). The

**Table 2** Dose reduction potential of the DLM and DLR for 10 mm, 5 mm, and 1 mm target objects

	10 mm	5 mm	1 mm	Average	Min	Max
DLM	87	87	84	86	81	88
DLR-H	86	89	88	87	78	92
DLR-M	74	78	76	76	60	81
DLR-L	58	61	61	60	46	67

*FBP* filtered back projection, *DLM* deep learning model, *DLR-H* deep learning-based image reconstruction algorithm, high strength, *DLR-M* deep learning-based image reconstruction algorithm, medium strength, *DLR-L* deep learning-based image reconstruction algorithm, low strength

CT values of the five inlets and the background were relatively constant across the radiation doses and phantom sizes.

There was a tendency where the DLM showed higher spatial resolution than the DLRs for inlets having high contrast values (bone-mimicking material and iodine) (Table 3). On the other hand, in the case of polystyrene, being the inlet with the lowest contrast value among the five inlets, the spatial resolution of DLM was degraded compared with the FBP; that of the DLM was between that of the medium and high reconstruction levels of the DLR. For air, the spatial resolution of the DLR was higher than that of the DLM.

Moreover, the spatial resolutions of the DLM and DLR were dependent on the level of radiation exposure (Table 4).  $TTF_{50\%}$  of FBP, DLM, and DLR-H, M, and L decreased as the radiation dose decreased. The spatial resolution of DLR decreased with the increase of reconstruction strengths.

## Discussion

In this study, performances of the vendor-agnostic DLM and a vendor-specific DLR were compared in terms of the DRP. We found that the DLM showed DRP comparable to that of DLR. Although several vendor-specific DLRs are now available, to the knowledge of the authors, this study is the first to compare a vendor-agnostic DLM with a commercially available DLR based on the method specified by the AAPM TG 233 group.

Analysis of CT image quality based on conventional metrics such as CNR or modulation transfer function is insufficient for an objective evaluation of DL algorithms, as those metrics cannot appropriately evaluate the non-linear and multifarious nature of the DL algorithms [24]. To reflect the

**Table 3** A percentage change (range) of a spatial resolution (TTF<sub>50%</sub>) of the DLM and DLR compared to FBP according to inlets

	Iodine (335 HU)	Polystyrene (50 HU)	Bone (1000 HU)	Water (90 HU)	Air (-895 HU)
DLM	5 (15 to -1)	-28 (-41 to -13)	3 (-9 to 13)	-6 (-24 to 15)	0 (-8 to 6)
DLR-H	-2 (-9 to 2)	-31 (-39 to -24)	-3 (-9 to 1)	-11 (-18 to -4)	3 (-2 to 6)
DLR-M	0 (-7 to 4)	-25 (-33 to -17)	-2 (-9 to 2)	-8 (-16 to -2)	3 (-2 to 6)
DLR-L	2 (-5 to 5)	-19 (-25 to -13)	-2 (-8 to 2)	-5 (-13 to 0)	3 (5 to -3)

TTF task transfer function, FBP filtered back projection, DLM deep learning model, DLR-H deep learning-based image reconstruction algorithm, high strength, DLR-M deep learning-based image reconstruction algorithm, medium strength, DLR-L deep learning-based image reconstruction algorithm, low strength

object-dependent resolution and noise performance of DL algorithms, we employed metrics (e.g., TTF, NPS, and d') by adopting the methodology specified by AAPM TG 233 to better represent the clinical performance.

It was unclear whether the vendor-agnostic DLM could exhibit DRP performance comparable to that of the DLR. The DLR was expected to have integrated into its neural networks many vendor-specific experiences and knowledge regarding the physical and mathematical properties of the whole reconstruction process. In contrast, the DLM focuses on a single task of image noise reduction without using raw projection data.

Therefore, our finding that the DLM showed DRP comparable to that of the DLR is encouraging for attempts to reduce radiation doses. The algorithm can potentially be adapted by many institutions having CT scanners from different vendors.

Indeed, compared to reported DRPs of the currently used IRs (range from 25 to 65%) (4, 5), the DLM in our study showed DRP of a higher range (81–88%). If this degree of dose reduction potential can be translated into clinical practice, CT images of the abdomen could be acquired under a 2-mGy radiation dose (15 mGy × 14%). When the effective dose (mSv) is calculated with a constant 0.015 mSv·mGy<sup>-1</sup>·cm<sup>-1</sup> and a scan length of 50 cm, the radiation exposure is approximately 1.6 mSv, which is equivalent to that of two abdomen x-rays [25]. Previous studies have shown that clinical outcome using 2-mSv CT is not inferior to that using conventional dose CT for acute appendicitis. There is a dearth

of similar research in other disease groups [26, 27]. Our results may facilitate future research on reducing CT radiation used for various diseases

In the subgroup analysis, the DLM showed d' higher than that of the DLR as the diameter of the phantom increased. Such strength of DLM may be translated to its application in obese patients. In low radiation-dose levels, however, the DLM showed lower d' than did the DLR-H, owing to its relatively low noise reduction. This performance change may be attributed to the difference in denoising principles between the two algorithms. Because the DLM is an image-based noise reduction algorithm, it was trained to reduce the noise level only to the extent where the image texture could be preserved. In contrast, in the DLR, which is a projection sinogram-based algorithm, the noise reduction power might override the preservation of image texture. In fact, a previous study reported that when applied to low radiation dose CT scans in its full strength, the DLR caused image texture degradation and texture distortion artifacts that can be mistaken for a true lesion [15].

In previous DLR studies, the image texture was preserved, and the NPS peak remained unchanged compared with the FBP [28]. However, analysis of the noise and spatial resolution revealed that the DLM showed a slightly right-shifted NPS peak, reflecting a sharpening of the image texture. This shift of the NPS peak in the DLM resulted in higher d' values than did the DLR-H for all object materials. Further evaluation is needed to

**Table 4** Mean TTF<sub>50%</sub> values of the FBP, DLM, and DLR according to radiation dose levels

	15.47 mGy	7.74 mGy	3.87 mGy	1.93 mGy	0.97 mGy	0.48 mGy
FBP	0.41 (-)	0.40 (-)	0.38 (-)	0.36 (-)	0.33 (-)	0.29 (-)
DLM	0.44 (6)	0.42 (5)	0.40 (4)	0.29 (-18)	0.27 (-18)	0.25 (-15)
DLR-H	0.39 (-5)	0.38 (-5)	0.36 (-7)	0.33 (-9)	0.29 (-12)	0.26 (-13)
DLR-M	0.40 (-3)	0.39 (-3)	0.37 (-4)	0.34 (-6)	0.30 (-10)	0.26 (-10)
DLR-L	0.41 (-1)	0.39 (-1)	0.37 (-2)	0.35 (-3)	0.31 (-7)	0.27 (-8)

TTF task transfer function, FBP filtered back projection, DLM deep learning model, DLR-H deep learning-based image reconstruction algorithm, high strength, DLR-M deep learning-based image reconstruction algorithm, medium strength, DLR-L deep learning-based image reconstruction algorithm, low strength

The numbers in the parentheses indicate the percentage change of mean spatial resolution of DLM and DLR compared to FBP for the same radiation dose levels

determine whether such shift of the NPS peak of the DLM affects clinical performance.

Previous studies have demonstrated that for DLR, spatial resolution depends on contrast of the object material and radiation dose [29–31]. Our results were consistent with those of the previous studies, showing a reduction in the spatial resolution (TTF) with a lower radiation dose and target contrast, in both DLR and DLM images. Owing to such dependency, McCollough et al found that a radiation dose reduction of more than 25% results in lower spatial resolution in IR images than in FBP images [30]. Therefore, the decrease in spatial resolution is likely to be worse for objects having contrast lower than that of polystyrene (50 HU). To overcome the degradation of the spatial resolution, the use of a generative adversarial network (GAN) was proposed in deep learning–based algorithms. The GAN focuses on migrating the data noise distribution from the strong to the weak statistically, to avoid over-smoothing or structural detail losses [32]. However, verification of GAN in clinical CT remains insufficient. Therefore, further study is needed to determine whether the DLM incorporating a GAN can further preserve the spatial resolution for low-contrast object materials. Meanwhile, our results also showed the degradation of spatial resolution of DLM below 1.97 mGy. There is no clear explanation for this phenomenon, and further research is needed regarding how the spatial resolution of DLM is affected at a very low radiation dose level.

This study has several limitations. First, because the DLM and DLR were evaluated using a phantom, further research using clinical CT images is needed to validate their performance in clinical applications. Second, although we tried to remove vendor-specific overfitting bias by training with various combinations of clinical and phantom images from four vendors, the DLM and DLR were evaluated using only one type of CT equipment. Therefore, to demonstrate the generalizability of the DLM, it should be evaluated on various other CT machines. Third, because we used FBP images reconstructed only with the body kernel, further research using other kernels may be needed. Finally, as mentioned, owing to the dependency of spatial resolution on contrast, further study is needed on low-contrast objects (below 50 HU) based on the AAPM TG 233 guideline to validate the clinical applicability of the DLM.

In conclusion, the dose reduction potential of the vendor-agnostic DLM was shown to be comparable to that of the vendor-specific DLR at high strength and superior to those of the DLRs at medium and low strengths

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-021-08199-9>.

**Funding** This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Korean government (\*MSIT) (grant number NRF-2018R1C1B6007999).

## Declarations

**Guarantor** The scientific guarantor of this publication is Won Chang.

**Conflict of interest** ClariPI provided technical support in obtaining denoised images. Jong Hyo Kim is CO-CEO & CTO of ClariPI but did not have control over any of the data or information submitted for publication.

**Statistics and biometry** No complex statistical methods were necessary for this paper.

**Informed consent** Written informed consent was waived by the Institutional Review Board.

**Ethical approval** Institutional Review Board approval was obtained

## Methodology

- Experimental
- Performed at one institution

## References

1. Brenner DJ, Hall EJ (2007) Computed tomography—an increasing source of radiation exposure. *N Engl J Med* 357:2277–2284
2. Smith-Bindman R, Lipson J, Marcus R et al (2009) Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Arch Intern Med* 169:2078–2086
3. Hara AK, Paden RG, Silva AC, Kujak JL, Lawder HJ, Pavlicek W (2009) Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. *AJR Am J Roentgenol* 193:764–771
4. Leipsic J, LaBounty TM, Heilbron B et al (2010) Estimated radiation dose reduction using adaptive statistical iterative reconstruction in coronary CT angiography: the ERASIR study. *AJR Am J Roentgenol* 195:655–660
5. Katsura M, Matsuda I, Akahane M et al (2012) Model-based iterative reconstruction technique for radiation dose reduction in chest CT: comparison with the adaptive statistical iterative reconstruction technique. *Eur Radiol* 22:1613–1623
6. 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
7. Tatsugami F, Higaki T, Nakamura Y et al (2019) Deep learning–based image restoration algorithm for coronary CT angiography. *Eur Radiol* 29:5322–5329
8. 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
9. 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
10. Kim JH, Yoon HJ, Lee E, Kim I, Cha YK, Bak SH (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
11. Homayounieh F, Holmberg O, Al Umairi R et al (2021) Variations in CT utilization, protocols, and radiation doses in COVID-19 pneumonia: results from 28 countries in the IAEA study. *Radiology* 298:e141–e151
  12. Lim WH, Choi YH, Park JE et al (2019) Application of vendor-neutral iterative reconstruction technique to pediatric abdominal computed tomography. *Korean J Radiol* 20:1358–1367
  13. Hong JH, Park E-A, Lee W, Ahn C, Kim J-H (2020) Incremental image noise reduction in coronary CT angiography using a deep learning-based technique with iterative reconstruction. *Korean J Radiol* 21:1165–1177
  14. Kolb M, Storz C, Kim JH et al (2019) Effect of a novel denoising technique on image quality and diagnostic accuracy in low-dose CT in patients with suspected appendicitis. *Eur J Radiol* 116:198–204
  15. Nam JG, Ahn C, Choi H et al (2021) Image quality of ultralow-dose chest CT using deep learning techniques: potential superiority of vendor-agnostic post-processing over vendor-specific techniques. *Eur Radiol*. <https://doi.org/10.1007/s00330-020-07537-7>
  16. Samei E, Bakalyar D, Boedeker KL et al (2019) Performance evaluation of computed tomography systems: summary of AAPM task group 233. *Med Phys* 46:e735–e756
  17. Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL (2017) US diagnostic reference levels and achievable doses for 10 adult CT examinations. *Radiology* 284:120–133
  18. 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
  19. 510(k) Premarket Notification (device name ClariCT.AI) (2019) FDA. Available via [https://www.accessdata.fda.gov/cdrh\\_docs/pdf18/K183460.pdf](https://www.accessdata.fda.gov/cdrh_docs/pdf18/K183460.pdf). Accessed 03 May
  20. Ahn C, Heo C, Kim JH (2019) Combined low-dose simulation and deep learning for CT denoising: application in ultra-low-dose chest CT, International Forum on Medical Imaging in Asia 2019
  21. Lee S, Choi YH, Cho YJ et al (2020) Noise reduction approach in pediatric abdominal CT combining deep learning and dual-energy technique. *Eur Radiol*. <https://doi.org/10.1007/s00330-020-07349-9>
  22. Greffier J, Frandon J, Larbi A, Beregi JP, Pereira F (2020) CT iterative reconstruction algorithms: a task-based image quality assessment. *Eur Radiol* 30:487–500
  23. Samei E, Richard S (2015) Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology. *Med Phys* 42:314–323
  24. Christianson O, Chen JJ, Yang Z et al (2015) An improved index of image quality for task-based performance of CT iterative reconstruction across three commercial implementations. *Radiology* 275:725–734
  25. Mettler FA Jr, Huda W, Yoshizumi TT, Mahesh M (2008) Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology* 248:254–263
  26. Kim HJ, Jeon BG, Hong CK et al (2017) Low-dose CT for the diagnosis of appendicitis in adolescents and young adults (LOCAT): a pragmatic, multicentre, randomised controlled non-inferiority trial. *Lancet Gastroenterol Hepatol* 2:793–804
  27. Kim K, Kim YH, Kim SY et al (2012) Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med* 366:1596–1605
  28. 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
  29. Yu L, Vrieze TJ, Leng S, Fletcher JG, McCollough CH (2015) Measuring contrast-and noise-dependent spatial resolution of an iterative reconstruction method in CT using ensemble averaging. *Med Phys* 42:2261–2267
  30. 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
  31. Li K, Garrett J, Ge Y, Chen GH (2014) Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance. *Med Phys* 41:071911
  32. Yang Q, Yan P, Zhang Y et al (2018) Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging* 37:1348–1357

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