Combination of Deep Learning–Based Denoising and Iterative Reconstruction for Ultra-Low-Dose CT of the Chest: Image Quality and Lung-RADS Evaluation

Akinori Hata[1]
Masahiro Yanagawa[2]
Yuriko Yoshida[2]
Tomo Miyata[2]
Mitsuko Tsubamoto[1]
Osamu Honda[3]
Noriyuki Tomiyama[2]

OBJECTIVE. The objective of our study was to assess the effect of the combination of deep learning–based denoising (DLD) and iterative reconstruction (IR) on image quality and Lung Imaging Reporting and Data System (Lung-RADS) evaluation on chest ultra-low-dose CT (ULDCT).

MATERIALS AND METHODS. Forty-one patients with 252 nodules were evaluated retrospectively. All patients underwent ULDCT (mean ± SD, 0.19 ± 0.01 mSv) and standard-dose CT (SDCT) (6.46 ± 2.28 mSv). ULDCT images were reconstructed using hybrid iterative reconstruction (HIR) and model-based iterative reconstruction (MBIR), and they were postprocessed using DLD (i.e., HIR-DLD and MBIR-DLD). SDCT images were reconstructed using filtered back projection. Three independent radiologists subjectively evaluated HIR, HIR-DLD, MBIR, and MBIR-DLD images on a 5-point scale in terms of noise, streak artifact, nodule edge, clarity of small vessels, homogeneity of the normal lung parenchyma, and overall image quality. Two radiologists independently evaluated the nodules according to Lung-RADS using HIR, MBIR, HIR-DLD, and MBIR-DLD ULDCT images and SDCT images. The median scores for subjective analysis were analyzed using Wilcoxon signed rank test with Bonferroni correction. Intraobserver agreement for Lung-RADS category between ULDCT and SDCT was evaluated using the weighted kappa coefficient.

RESULTS. In the subjective analysis, ULDCT with DLD showed significantly better scores than did ULDCT without DLD ($p < 0.001$), and MBIR-DLD showed the best scores among the ULDCT images ($p < 0.001$) for all items. In the Lung-RADS evaluation, HIR showed fair or moderate agreement (reader 1 and reader 2: κw = 0.46 and 0.32, respectively); MBIR, moderate or good agreement (κw = 0.68 and 0.57); HIR-DLD, moderate agreement (κw = 0.53 and 0.48); and MBIR-DLD, good agreement (κw = 0.70 and 0.72).

CONCLUSION. DLD improved the image quality of both HIR and MBIR on ULDCT. MBIR-DLD was superior to HIR_DLD for image quality and for Lung-RADS evaluation.

Keywords: artificial intelligence (AI), computer-assisted image processing, image processing, image reconstruction, lung neoplasms, MDCT, radiation dosage

doi.org/10.2214/AJR.19.22680

Received December 10, 2019; accepted after revision March 7, 2020.

Supported by technical support from AlgoMedica, which also performed image processing using PixelShine. AlgoMedica had no role in study design, analysis, or manuscript preparation.

[1]Department of Future Diagnostic Radiology, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan. Address correspondence to A. Hata (a-hata@radiol.med.osaka-u.ac.jp).

[2]Department of Diagnostic and Interventional Radiology, Osaka University Graduate School of Medicine, Osaka, Japan.

[3]Department of Radiology, Kansai Medical University, Osaka, Japan.

AJR 2020; 215:1321–1328

ISSN-L 0361–803X/20/2156–1321

As CT has become common in medical practice, radiation exposure has increased markedly. There is a worldwide need to optimize the radiation dose to the patient according to the principle of as low as reasonably achievable (ALARA) [1]. In conventional filtered back projection (FBP) reconstruction of CT images, there is a trade-off between image quality and radiation dose; however, recent iterative reconstruction (IR) techniques have enabled a significant decrease in radiation dose without degrading image quality [2–12]. Investigators have reported that IR images from ultra-low-dose CT (ULDCT), for which the radiation dose is equivalent to a chest radiograph, are feasible for several purposes [13–18].

Hybrid iterative reconstruction (HIR) and model-based iterative reconstruction (MBIR) are available clinically. MBIR is a pure IR algorithm that considers system statistics and system optics and is more mathematically complex than HIR. It is generally accepted that compared with CT images with FBP and HIR, CT images with MBIR have less noise, improved spatial resolution, and increased contrast resolution [2, 7]. However, MBIR requires an additional vendor-specific reconstruction unit and a lengthy reconstruction time, which may interfere with clinical practice.

Artificial intelligence technologies such as deep learning are recently emerged techniques that are being applied across various areas [19–21]. The deep learning technique

is suitable for image postprocessing to decrease noise and enhance image quality. Investigators have reported that deep learning–based denoising (DLD) significantly reduces image noise and improves the signal-to-noise and contrast-to-noise ratios on pelvic CT [22]. DLD is developed using huge amounts of high-dose and low-dose data, a high-performance graphics processing unit (GPU), and a long training time. After development, a DLD technique can then be implemented on a computer without a GPU and can process images rapidly. Because the processing is image-based, it does not require raw CT data. In contrast, IR is a raw data–based technique. Although these two techniques are quite different, they can be used simultaneously. However, no studies, to our knowledge, have reported the application of DLD in combination with IR in ULDCT for clinical use. In particular, it is not known which technique is more advantageous for the preprocessing base image: MBIR or HIR. If HIR with DLD (HIR-DLD) shows good enough performance, HIR-DLD will be more useful in clinical practice than MBIR with DLD (MBIR-DLD) because of the short reconstruction and image-processing time.

Lung cancer screening can be given as an example of when the IR and DLD technology may be helpful. The National Lung Screening Trial found that low-dose CT screening reduces mortality from lung cancer [23], which is a major cause of death internationally. Accordingly, demand for lung cancer screening CT at a low dose is increasing. The American College of Radiology has published a quality assurance tool termed the Lung Imaging Reporting and Data System (Lung-RADS). Lung-RADS is designed to standardize the reporting of CT screening for lung cancer and management recommendations and to facilitate outcome monitoring [24]. Lung-RADS uses categories to differentiate high-risk nodules from low-risk nodules according to nodule type, nodule size, and growth. It is challenging to detect nodules with subsolid (i.e., part solid and nonsolid) types clearly on CT images obtained at a low radiation dose because of their low contrast and small structure.

The purpose of this study was to assess the effect of the combination of DLD and IR on image quality and Lung-RADS evaluation on chest ULDCT.

Materials and Methods

Patients

This retrospective study was approved by the institutional review board at Osaka University Graduate School of Medicine. All patients gave written informed consent for the use of their CT data for research. Standard-dose CT (SDCT) and ULDCT data were available for 41 patients who were scanned between January 2016 and February 2017. The CT scans were obtained for follow-up of pulmonary subsolid nodules. The ULDCT scans were originally acquired for other research regarding image quality with MBIR. The age and body mass index (BMI [weight in kilograms divided by the square of height in meters]) of each patient were recorded. Data from 21 patients have been previously reported in a separate study that evaluated the effect of IRs on emphysema quantification [17].

CT Acquisition and Image Reconstruction

All CT images were obtained using an MDCT scanner (Discovery CT750HD, GE Healthcare). Unenhanced CT was performed at a standard dose and at an ultra low dose. The SDCT scan was obtained first, followed immediately by an ULDCT scan obtained in a separate breath-hold. SDCT was performed using a routine clinical protocol with automatic tube current modulation. ULDCT was performed with a tube current of 10 mA. Otherwise, the protocols for SDCT and ULDCT were identical: tube voltage, 120 kVp; gantry rotation period, 0.4 second; detector collimation, 0.625 mm; detector pitch, 0.984; and non–high-resolution mode with 984 views per rotation. All examinations were performed with the patient in the supine position at end-inspiration. The volume CT dose index ($CTDI_{vol}$) and dose-length product (DLP) were recorded for all CT examinations, and effective dose (ED) was calculated using a conversion coefficient (0.014 mSv/mGy × cm) for chest CT [25].

ULDCT was reconstructed using HIR (Adaptive Statistical Iterative Reconstruction [ASiR], GE Healthcare) and MBIR (Veo 3.0, GE Healthcare) with a thickness of 0.625 mm, matrix size of 512 × 512, and FOV of 34.5 × 34.5 cm. To enable preliminary evaluation, the ULDCT scans of two patients were reconstructed using FBP with a sharp kernel; HIR with blending settings of 30%, 60%, and 100%; and MBIR with lung (RP20) and standard (Stnd) settings. The RP20 setting is used specifically to provide high-resolution images for lung evaluation, whereas the Stnd (normal) setting is used for imaging the head, mediastinum, and abdomen [26].

The reconstructed images were processed with commercially available software using a DLD algorithm (PixelShine, AlgoMedica). The principal investigator (7 years of radiology experience) evaluated the images and chose HIR with the 30% blending setting and MBIR with the RP20 setting for further evaluations because the postprocessed FBP and HIR images of various settings were similar to each other and the postprocessed MBIR images with RP20 and Stnd settings were similar to each other.

The ULDCT scans of all patients were reconstructed using HIR with 30% blending setting (HIR set) and MBIR with RP20 (MBIR set). Both image sets were postprocessed using the DLD algorithm (HIR-DLD set and MBIR-DLD set). SDCT scans were reconstructed using FBP with a sharp kernel (SDCT set). ULDCT and SDCT images were reconstructed with a thickness of 0.625 mm, matrix size of 512 × 512, and FOV of 34.5 × 34.5 cm. In total, we obtained five image sets: HIR, MBIR, HIR-DLD, MBIR-DLD, and SDCT.

Subjective Image Analysis

The principal investigator reviewed the SDCT image sets and identified the largest subsolid nodule for each patient. Images at the slice showing the maximum diameter of the nodule were extracted. ULDCT images at the same slice level were also obtained. In total, 205 evaluation images (5 reconstructions per patient for 41 patients) were obtained for subjective analysis. In addition, an anthropomorphic thoracic CT phantom (N1 Lungman, Kyoto Kagaku) containing a ground-glass nodule was scanned on the same CT unit, and these scans were used as the reference standard images. Three scans were obtained at different tube current settings: 10 mA (minimum tube current), 75 mA (30 mAs, which is common for low-dose CT [LDCT]), and 390 mA (maximum tube current). All other CT protocols were the same as for the patient scans. Images were reconstructed using FBP with a sharp kernel. The slices at the level of the carina were used as the reference standard.

Three chest radiologists (27, 18, and 4 years of radiology experience, respectively) independently and subjectively analyzed the evaluation images on a 3-megapixel 21-inch (53-cm) monochrome liquid crystal display monitor. All images were displayed at a window level of –600 HU and a window width of 1500 HU.

The observers were blinded to the image reconstruction type and the CT protocol used for each image, and they compared each evaluation image with the reference standard images. Noise, streak artifact, clarity of small vessels, nodule edge, homogeneity of normal lung parenchyma, and overall image quality were graded using the following 5-point scale: 5, equivalent or superior to the 390-mA reference standard image; 4, inferior to a score of 5 but superior to a score of 3; 3, equivalent to the 75-mA reference standard image; 2, inferior to a score of 3 but superior to a score of 1; and 1, equivalent or inferior to the 10-mA reference standard image. The areas in which the radiologists

should evaluate small vessels, nodule edge, and homogeneity of normal lung parenchyma were indicated on a hard-copy SDCT image.

Lung-RADS Evaluation

The principal investigator reviewed the SDCT image set and identified all nodules, including solid tiny nodules. Two chest radiologists (readers 1 and 2 with 23 and 3 years of radiology experience, respectively) independently evaluated the nodules on the HIR, MBIR, HIR-DLD, MBIR-DLD, and SDCT image sets. The two readers were blinded to the reconstruction techniques. The SDCT images were printed on paper, and the paper was used to indicate the location of the nodules to the readers. In the images on the paper, the nodules were masked because it would affect the readers' decision.

Two radiologists categorized the nodules according to Lung-RADS version 1.0 criteria [23]. All image sets were considered as baseline screening scans. We did not use category 4X in our study because it is based on subjective morphologic criteria other than nodule type and size [27]. Nodules that could not be identified in the evaluation images were categorized as 1. The nodule type (i.e., solid, part-solid, nonsolid, or unidentified) was also recorded. After evaluation by the readers 1 and 2, the principal investigator reviewed the SDCT results. If there was any discrepancy in results between readers 1 and 2, the principal investigator reviewed and categorized the SDCT image again, and the final result was considered the reference standard. To evaluate nodule size, the principal investigator measured both the long-axis and short-axis dimensions of the nodule on SDCT and calculated the mean diameter of each nodule.

Statistical Analysis

All statistical analyses were performed using R software (version 3.4.1, R Foundation for Statistical Computing). In the subjective analysis, the median values of the three radiologists' scores were analyzed statistically. Differences in scores among the four reconstructions were tested using the Wilcoxon signed rank test, which was conducted with a Bonferroni correction for multiple comparisons.

For the Lung-RADS evaluation, intraobserver agreement of Lung-RADS categories between SDCT and ULDCT was assessed with a weighted kappa coefficient (κw) for each radiologist. Interobserver agreement was also evaluated for each reconstruction method. Regarding nodules that were not identified on ULDCT, the nodule type and Lung-RADS category on the reference standard were assessed to identify the characteristics of those nodules. The statistical significance of the number of unidentified nodules on ULDCT was assessed using the McNemar test.

To evaluate the detectability of significant nodules, sensitivity and specificity were evaluated for each reconstruction method, and nodules of category 3 or higher on the reference standard were considered positive [23]. An ROC curve analysis was used to determine area under the ROC curves (AUC) of each reconstruction. The two AUCs were compared with each other by the DeLong method [28].

A p value of < 0.05 and a Bonferroni-corrected p value of < 0.0083 (0.05 / 6) were considered significant. We categorized the weighted kappa as follows: poor ($0 < \kappa w \leq 0.20$), fair ($0.20 < \kappa w \leq 0.40$), moderate ($0.40 < \kappa w \leq 0.60$), good ($0.60 < \kappa w \leq 0.80$), and excellent ($0.80 < \kappa w \leq 1.00$) [29].

Results

Patient Demographic and Clinical Characteristics and Radiation Dose

The patient demographic characteristics were as follows: 17 men and 24 women with a mean age of 68.5 years (age range, 42–83 years). The mean BMI was 21.7 (range, 17.4–31.6). A total of 252 nodules were identified in 41 patients on SDCT.

For SDCT, mean $CTDI_{vol}$, mean DLP, and mean ED were 11.14 ± 3.63 (SD) mGy, 461.19 ± 162.88 mGy × cm, and 6.46 ± 2.28 mSv, respectively. For ULDCT, mean $CTDI_{vol}$, mean DLP, and mean ED were 0.32 ± 0.00 mGy, 13.22 ± 0.92 mGy × cm, and 0.19 ± 0.01 mSv.

Subjective Image Analysis

Table 1 lists the mean scores of the subjective analyses. For all items, HIR-DLD scored significantly better than HIR, and MBIR-DLD scored significantly better than MBIR ($p < 0.001$). MBIR-DLD scored the best among the ULDCT images for all items ($p < 0.001$; Fig. 1). Regarding overall image quality, the mean score of MBIR-DLD was

TABLE 1: Comparison of Subjective Scores Among Ultra-Low-Dose CT (ULDCT) Reconstruction Techniques

ULDCT Reconstruction	Noise	Streak Artifact	Clarity of Small Vessels	Nodule Edge	Homogeneity of Normal Lung Parenchyma	Overall Image Quality
Score[a] (mean ± SD)						
HIR	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.2	1.0 ± 0.0	1.0 ± 0.0
MBIR	4.2 ± 0.5	4.3 ± 0.6	2.4 ± 0.5	2.6 ± 0.7	2.0 ± 0.2	2.6 ± 0.6
HIR-DLD	3.4 ± 0.5	2.3 ± 0.6	2.0 ± 0.6	2.4 ± 0.9	2.7 ± 0.6	2.5 ± 0.6
MBIR-DLD	4.9 ± 0.3	5.0 ± 0.0	3.3 ± 0.5	3.5 ± 0.7	3.6 ± 0.5	4.0 ± 0.6
p[b]						
HIR vs MBIR	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]
HIR vs HIR-DLD	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]
HIR vs MBIR-DLD	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]
MBIR vs HIR-DLD	< 0.001[c]	< 0.001[c]	< 0.001[c]	0.218	< 0.001[c]	0.385
MBIR vs MBIR-DLD	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]
HIR-DLD vs MBIR-DLD	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]	< 0.001[c]

Note—HIR = hybrid iterative reconstruction, MBIR = model-based iterative reconstruction, HIR-DLD = HIR with deep learning–based denoising (DLD), MBIR-DLD = MBIR with DLD.

[a]The following 5-point scale was used: 5, equivalent or superior to the 390-mA reference standard image; 4, inferior to a score of 5 but superior to a score of 3; 3, equivalent to the 75-mA reference standard image; 2, inferior to a score of 3 but superior to a score of 1; and 1, equivalent or inferior to the 10-mA reference standard image.

[b]The differences in scores among the reconstructions were assessed using Wilcoxon signed rank test.

[c]Difference is statistically significant. A Bonferroni-corrected p value of < 0.0083 (0.05 / 6) was considered significant.

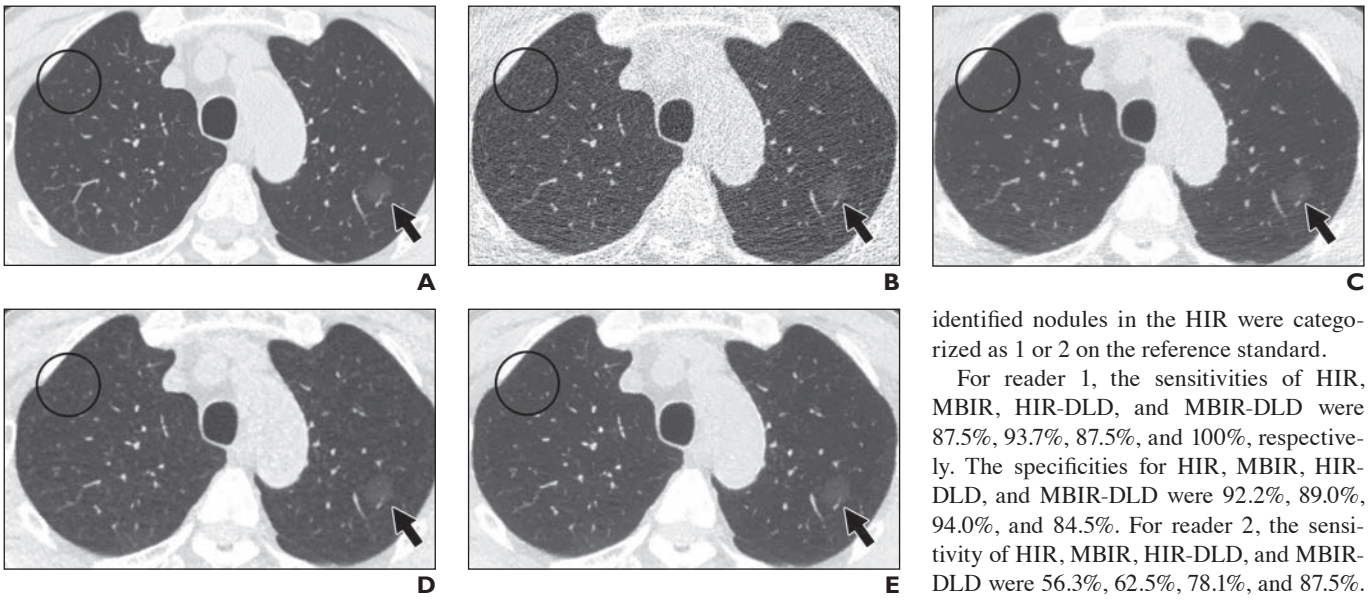

Fig. 1—75-year-old man with nonsolid nodule. Body mass index is 20.8.
A–E, Axial standard-dose CT (SDCT) image (**A**) and axial ultra-low-dose CT (ULDCT) images reconstructed using hybrid iterative reconstruction (HIR) (**B**), HIR with deep learning–based denoising (DLD) (HIR-DLD) (**C**), model-based iterative reconstruction (MBIR) (**D**), and MBIR with DLD (MBIR-DLD) (**E**). Severe noise is apparent in HIR (**B**) and is significantly improved in HIR-DLD (**C**), MBIR (**D**), and MBIR-DLD (**E**). Streak artifact is present in HIR (**B**) and HIR-DLD (**C**) and affects visibility of nodule (*arrows*). However, streak artifact is not seen in MBIR images (**D** and **E**). Circles show normal lung area without emphysema. In SDCT image (**A**), this area appears normal (*circle*, **A**). In MBIR image (**D**), this area shows inhomogeneity suggestive of emphysema (*circle*, **D**). Although this area (*circle*, **C**) is homogeneous in HIR-DLD image (**C**), visibility of small vessels is degraded. In MBIR-DLD image (**E**), appearance of area (*circle*, **E**) seen as artificial inhomogeneity in MBIR image (**D**) is improved without reduction in visibility of small vessels. MBIR-DLD shows best overall image quality among ULDCT images.

4.0, indicating an overall image quality superior to the 75-mA (LDCT) reference standard image. Scores of noise, streak artifact, and the clarity of small vessels for HIR-DLD were significantly better than those for HIR but were worse than those for MBIR ($p < 0.001$). The homogeneity of normal lung parenchyma for HIR-DLD was better than for MBIR ($p < 0.001$; Fig. 2). Regarding nodule edge and overall image quality, there was no significant difference between MBIR and HIR-DLD.

Lung-RADS Evaluation

Nodule characteristics on the reference standard, the results of intraobserver agreement and interobserver agreement, and characteristics of nodules classified as unidentified on ULDCT are summarized in Tables 2, 3, and 4, respectively.

Intraobserver agreements between HIR and SDCT were fair or moderate (reader 1 and reader 2: κw = 0.46 and 0.32, respectively); between MBIR and SDCT, moderate or good (κw = 0.68 and 0.57); between HIR-DLD and SDCT, moderate (κw = 0.53 and 0.48); and between MBIR-DLD and SDCT, good (κw = 0.70 and 0.72). The interobserver agreements between the two readers were moderate for HIR, MBIR, and MBIR-DLD and good for HIR-DLD and SDCT. For reader 1, 57.1% (144/252), 80.6% (203/252), 64.3% (162/252), and 79.8% (201/252) of the nodules showed the same category as SDCT on ULDCT images with HIR, MBIR, HIR-DLD, and MBIR-DLD, respectively. For reader 2, 49.2% (124/252), 79.4% (200/252), 67.1% (169/252), and 85.3% (215/252) of the nodules showed the same category as SDCT on ULDCT images with HIR, MBIR, HIR-DLD, and MBIR-DLD.

The number of unidentified nodules was significantly less for MBIR than for HIR ($p < 0.001$ for both readers; Fig. 3). HIR-DLD showed significantly fewer unidentified nodules than HIR ($p < 0.001$ for both readers). There was no significant difference between MBIR and MBIR-DLD ($p > 0.05$ for both readers). All nodules that were not identified in the MBIR, HIR-DLD, or MBIR-DLD images were categorized as 2 on the reference standard. One nodule (a part-solid nodule with a 9.8-mm diameter) that was not identified in the HIR for reader 2 was categorized as 3 on the reference standard. The other unidentified nodules in the HIR were categorized as 1 or 2 on the reference standard.

For reader 1, the sensitivities of HIR, MBIR, HIR-DLD, and MBIR-DLD were 87.5%, 93.7%, 87.5%, and 100%, respectively. The specificities for HIR, MBIR, HIR-DLD, and MBIR-DLD were 92.2%, 89.0%, 94.0%, and 84.5%. For reader 2, the sensitivity of HIR, MBIR, HIR-DLD, and MBIR-DLD were 56.3%, 62.5%, 78.1%, and 87.5%. The specificity of HIR, MBIR, HIR-DLD, and MBIR-DLD were 99.1%, 97.7%, 95.9%, and 95.9%. For reader 1, the AUCs for HIR, MBIR, HIR-DLD, and MBIR-DLD were 0.899, 0.914, 0.908, and 0.923. There was no significant difference in AUCs among the ULDCT reconstruction techniques ($p > 0.05$). For reader 2, the AUCs for HIR, MBIR, HIR-DLD, and MBIR-DLD were 0.777, 0.801, 0.870, and 0.917. MBIR-DLD showed significantly better performance than HIR ($p = 0.003$). The performance of MBIR-DLD tended to be better than that of MBIR, but the difference was not significant with Bonferroni correction ($p = 0.02$, which is > 0.0083). There was no significant difference among the other pairs of ULDCT images ($p > 0.05$).

Discussion

The results of our study show that DLD significantly improved the image quality of ULDCT with IR and that the image quality of MBIR with DLD was best among the four ULDCT techniques. In the Lung-RADS evaluation, MBIR-DLD on ULDCT showed better intraobserver agreement with SDCT and had fewer unidentified nodules than did HIR-DLD. Reader 2 showed improvement in performance values by use of DLD in both HIR and MBIR images. Reader 1 showed improvement in performance values by use of DLD in the number of unidentified nodules in HIR images, but the improvement in performance values by use of DLD was small in MBIR images.

The analysis of image quality revealed that noise was significantly reduced by DLD, es-

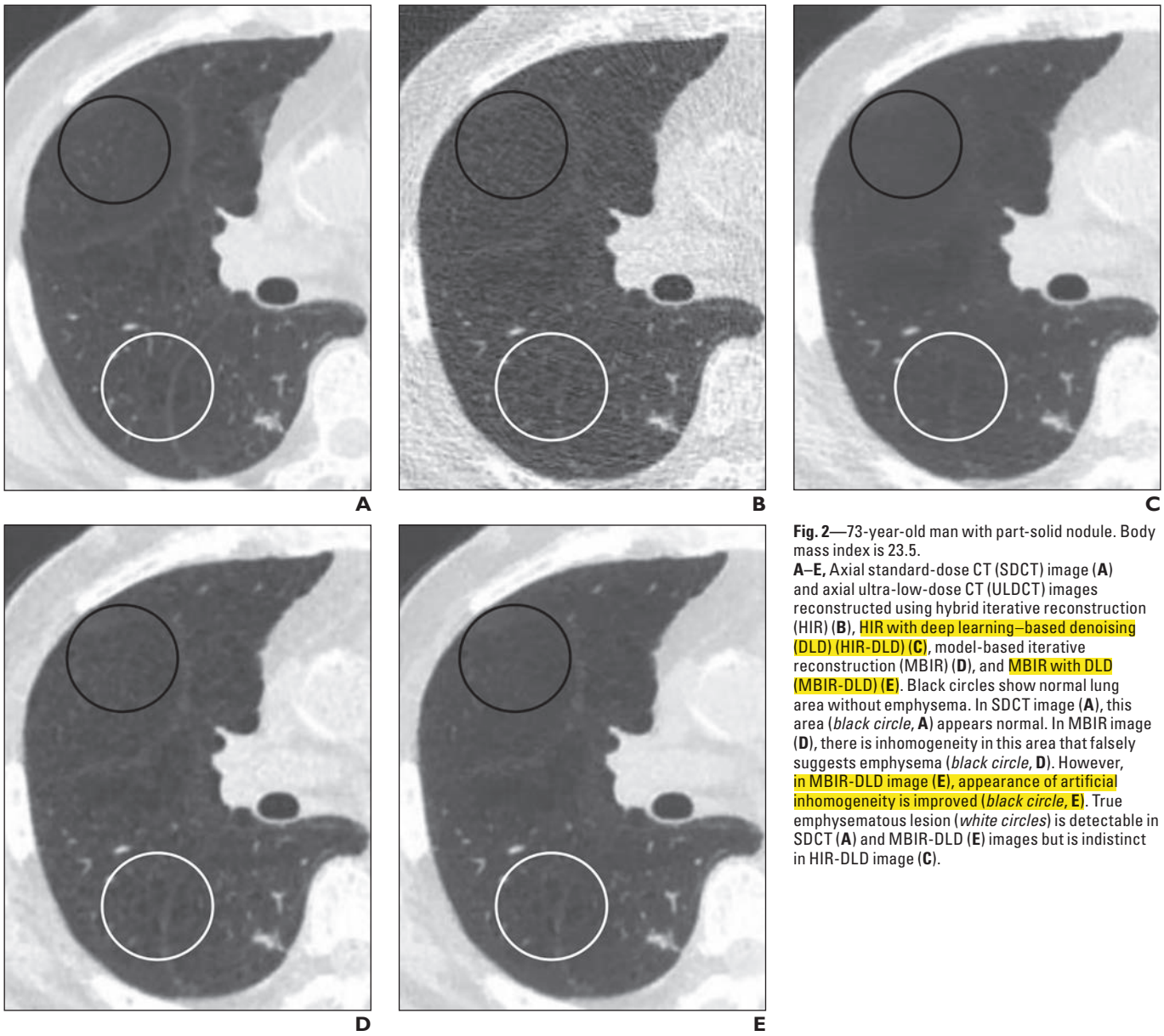

Fig. 2—73-year-old man with part-solid nodule. Body mass index is 23.5.
A–E, Axial standard-dose CT (SDCT) image (**A**) and axial ultra-low-dose CT (ULDCT) images reconstructed using hybrid iterative reconstruction (HIR) (**B**), HIR with deep learning–based denoising (DLD) (HIR-DLD) (**C**), model-based iterative reconstruction (MBIR) (**D**), and MBIR with DLD (MBIR-DLD) (**E**). Black circles show normal lung area without emphysema. In SDCT image (**A**), this area (*black circle*, **A**) appears normal. In MBIR image (**D**), there is inhomogeneity in this area that falsely suggests emphysema (*black circle*, **D**). However, in MBIR-DLD image (**E**), appearance of artificial inhomogeneity is improved (*black circle*, **E**). True emphysematous lesion (*white circles*) is detectable in SDCT (**A**) and MBIR-DLD (**E**) images but is indistinct in HIR-DLD image (**C**).

pecially noise in the HIR images. However, streak artifact was still present in the HIR-DLD images, and the mean score of HIR-DLD for streak artifact was significantly worse than those of the MBIR and MBIR-DLD images. DLD was developed mainly for denoising, not for streak artifact. Because the original ULDCT HIR images had extreme streak artifact, DLD was unable to remove them completely. In contrast, MBIR had little streak artifact on ULDCT, which may have contributed to better overall image quality.

In this study, artificial inhomogeneity was seen in areas of normal lung in the MBIR images without DLD. This artificial appearance is different from noise and could result in a false diagnosis of pulmonary emphysema. Although its cause is unclear, the artificial inhomogeneity may be related to the complex mathematic procedure of MBIR. DLD decreased this artificial inhomogeneity without reducing the visibility of structures such as small vessels. To obtain images of high resolution, we used MBIR with the lung setting rather than the standard setting. The artificial inhomogeneity could possibly be specific to MBIR with the lung setting; however, in our previous research, MBIR with the lung setting improved the visibility of small vessels and interstitial lung abnormalities in CT images compared with MBIR with the standard setting [18]. The deep learning system used in this study, which was trained using high-dose FBP CT images, enabled denoising and also enhanced the detail in the images. Because the image quality of MBIR-DLD was superior to MBIR without DLD for all items evaluated in this study, we speculate that DLD would contribute to improving the image quality of ULDCT images obtained using MBIR with the standard setting.

The positive effect of DLD on MBIR was relatively smaller than that on HIR in the image quality analysis. The image quality of the original HIR image on ULDCT was very low. On the other hand, it has been reported that MBIR significantly improved the image quality on ULDCT [3, 5]. This suggests that

TABLE 2: Nodule Characteristics on Reference Standard Images

Nodule Type, Lung-RADS Category	No. of Nodules	Size[a] (mm)
All nodules	252	4.6 (1.6–26.3)
1	12	
2	208	
3	17	
4A	11	
4B	4	
Solid nodules	64	3.0 (1.6–6.8)
1	12	
2	50	
3	2	
4A	0	
4B	0	
Part-solid nodules	57	6.8 (2.9–26.3)[b]
1	0	
2	27	
3	15	
4A	11	
4B	4	
Nonsolid nodules	131	5.1 (2.5–14.4)
1	0	
2	131	
3	0	
4A	0	
4B	0	

Note—Lung-RADS = Lung Imaging Reporting and Data System.
[a]Size is expressed as mean (range).
[b]Size of solid part: mean, 3.0 mm (range, 1.4–9.9 mm).

TABLE 3: Intraobserver and Interobserver Agreement of Two Readers for Lung-RADS Results

Imaging Technique	Intraobserver Agreement		Interobserver Agreement
	Reader 1	Reader 2	
HIR ULDCT vs SDCT	0.46	0.32	
MBIR ULDCT vs SDCT	0.68	0.57	
HIR-DLD ULDCT vs SDCT	0.53	0.48	
MBIR-DLD ULDCT vs SDCT	0.70	0.72	
HIR ULDCT			0.59
MBIR ULDCT			0.51
HIR-DLD ULDCT			0.61
MBIR-DLD ULDCT			0.56
SDCT			0.63

Note—Agreement was assessed using weighted kappa coefficient. Lung-RADS = Lung Imaging Reporting and Data System, HIR = hybrid iterative reconstruction, ULDCT = ultra-low-dose CT, SDCT = standard-dose CT, MBIR = model-based iterative reconstruction, HIR-DLD = HIR with deep learning–based denoising (DLD), MBIR-DLD = MBIR with DLD.

HIR images on ULDCT had greater room for improvement than MBIR images.

In the Lung-RADS evaluation, HIR-DLD showed worse intraobserver agreement with SDCT than did MBIR-DLD. The numerous unidentified nodules in HIR-DLD images had a negative effect on agreement. Because of the strong noise and streak artifact seen in the original HIR images, information of faint nodules was lost, and it was difficult to restore the image of the nodule even using DLD. However, all of the unidentified nodules in the HIR-DLD images were categorized as 2 on the reference standard, which suggests that the unidentified nodules may not be significant in the clinical setting.

The sensitivity for nodules of Lung-RADS category 3 or more was 88–100% for MBIR-DLD. Although the sensitivity of reader 2 was not 100% for MBIR-DLD, it also was not 100% on SDCT. The part-solid nodules were differentiated between categories 2 and 3 according to the presence (or not) of a solid component less than 6 mm, and several of the part-solid nodules were equivocal. Investigators have reported that interobserver agreement for the Lung-RADS category is substantial, but not perfect, for a weighted kappa coefficient of 0.67 (95% CI, 0.58–0.77) [27]. MBIR with DLD on ULDCT may be acceptable for screening; however, further evaluation is needed using a cohort with confirmed pathologic diagnoses.

At the ultra-low radiation dose, MBIR-DLD was advantageous in terms of image quality and Lung-RADS evaluation compared with HIR. The adoption of both DLD and MBIR will make a positive contribution to ULDCT for lung cancer screening. However, HIR has the advantage of a short reconstruction time, and MBIR is often unavailable (depending on the CT scanner used). DLD significantly improved the image quality of HIR, and the combination of DLD and HIR may be useful in many clinical situations at an appropriate radiation dose.

This study had several limitations. First, because most of the nodules in this study were not evaluated pathologically, we could not investigate the true diagnostic ability of the techniques for lung cancer. Second, to simplify the Lung-RADS evaluation, the nodule position was shown to the readers and category 4X nodules were discarded. Because the readers recorded only the category, we could not analyze the variability of the size measurement and size threshold for sensitivity. The

TABLE 4: Nodule Type on Reference Standard for Nodules That Were Classified as Unidentified on Ultra-Low-Dose CT (ULDCT)

Nodules	Reader 1				Reader 2			
	HIR	MBIR	HIR-DLD	MBIR-DLD	HIR	MBIR	HIR-DLD	MBIR-DLD
Unidentified nodules on ULDCT	69	5	35	4	102	18	53	13
Nodule type on reference standard								
Solid	9	1	6	0	25	4	12	5
Part-solid	4	0	0	0	4	0	2	0
Nonsolid	56	4	29	4	73	14	39	8

Note—HIR = hybrid iterative reconstruction, MBIR = model-based iterative reconstruction, HIR-DLD = HIR with deep learning–based denoising (DLD), MBIR-DLD = MBIR with DLD.

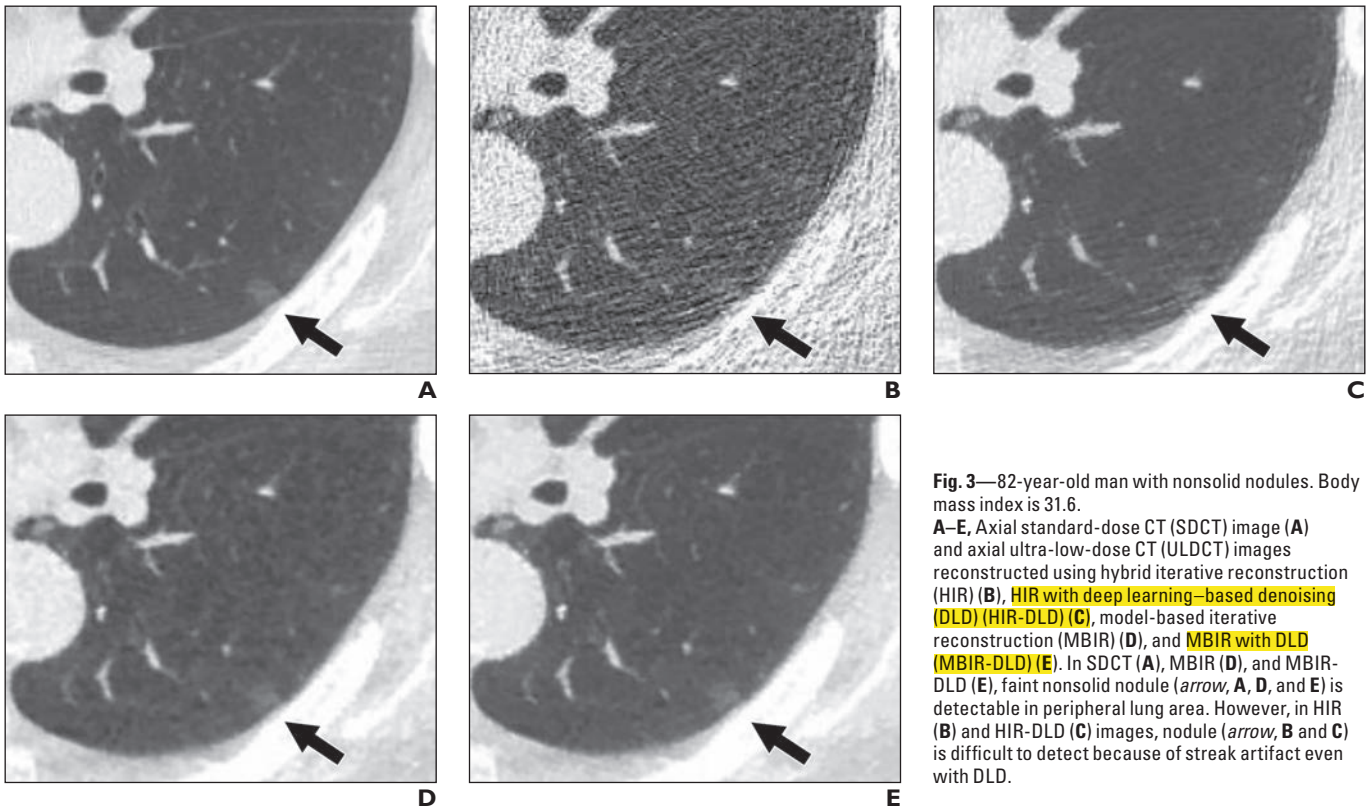

Fig. 3—82-year-old man with nonsolid nodules. Body mass index is 31.6.
A–E, Axial standard-dose CT (SDCT) image (**A**) and axial ultra-low-dose CT (ULDCT) images reconstructed using hybrid iterative reconstruction (HIR) (**B**), HIR with deep learning–based denoising (DLD) (HIR-DLD) (**C**), model-based iterative reconstruction (MBIR) (**D**), and MBIR with DLD (MBIR-DLD) (**E**). In SDCT (**A**), MBIR (**D**), and MBIR-DLD (**E**), faint nonsolid nodule (*arrow*, **A**, **D**, and **E**) is detectable in peripheral lung area. However, in HIR (**B**) and HIR-DLD (**C**) images, nodule (*arrow*, **B** and **C**) is difficult to detect because of streak artifact even with DLD.

detectability of nodules was not evaluated in this study. We plan further investigation of detectability using DLD on ULDCT. Third, the CT scanner, IR techniques, and DLD used were those of a single vendor. Fourth, we used an older version of Lung-RADS (version 1.0) because the latest version (version 1.1) [30] was not available at the time of the reading sessions. In addition, the growth rate of the nodules was not considered because only one ULDCT examination was available. Fifth, the solid nodule evaluation was limited because the study cohort had originally been obtained for the evaluation of subsolid nodules, and most of the solid nodules in this study were tiny and insignificant. Sixth, SDCT and ULDCT were analyzed in this study, but LDCT was not. However, recent CT lung cancer screenings are performed using LDCT. Comparison of ULDCT and LDCT would be useful to evaluate the performance of ULDCT with DLD. Seventh, SDCT was not evaluated in the subjective analysis. This would have added another valuable control. Finally, only one obese patient (BMI > 30) was included in the study. For scanning patients with a high BMI, a higher radiation dose is needed to obtain acceptable image quality, and different results may be obtained in a patient population with a high BMI.

In conclusion, DLD significantly improved the image quality of both HIR and MBIR images on ULDCT. MBIR was more advantageous than HIR in terms of image quality and Lung-RADS evaluation on ULDCT using DLD.

References

1. [No authors listed]. The ALARA (as low as reasonably achievable) concept in pediatric CT intelligent dose reduction: multidisciplinary conference organized by the Society of Pediatric Radiology—August 18–19, 2001. *Pediatr Radiol* 2002; 32:217–313
2. Katsura M, Matsuda I, Akahane M, et al. Model-

based iterative reconstruction technique for radiation dose reduction in chest CT: comparison with the adaptive statistical iterative reconstruction technique. *Eur Radiol* 2012; 22:1613–1623

3. Yamada Y, Jinzaki M, Tanami Y, et al. Model-based iterative reconstruction technique for ultralow-dose computed tomography of the lung: a pilot study. *Invest Radiol* 2012; 47:482–489
4. Pickhardt PJ, Lubner MG, Kim DH, et al. Abdominal CT with model-based iterative reconstruction (MBIR): initial results of a prospective trial comparing ultralow-dose with standard-dose imaging. *AJR* 2012; 199:1266–1274
5. Yanagawa M, Gyobu T, Leung AN, et al. Ultralow-dose CT of the lung: effect of iterative reconstruction techniques on image quality. *Acad Radiol* 2014; 21:695–703
6. Koc G, Courtier JL, Phelps A, Marcovici PA, MacKenzie JD. Computed tomography depiction of small pediatric vessels with model-based iterative reconstruction. *Pediatr Radiol* 2014; 44:787–794
7. Notohamiprodjo S, Deak Z, Meurer F, et al. Image quality of iterative reconstruction in cranial CT imaging: comparison of model-based iterative reconstruction (MBIR) and adaptive statistical iterative reconstruction (ASiR). *Eur Radiol* 2015; 25:140–146
8. Haggerty JE, Smith EA, Kunisaki SM, Dillman JR. CT imaging of congenital lung lesions: effect of iterative reconstruction on diagnostic performance and radiation dose. *Pediatr Radiol* 2015; 45:989–997
9. Annoni AD, Andreini D, Pontone G, et al. Ultralow-dose CT for left atrium and pulmonary veins imaging using new model-based iterative reconstruction algorithm. *Eur Heart J Cardiovasc Imaging* 2015; 16:1366–1373
10. Lubner MG, Pooler BD, Kitchin DR, et al. Submillisievert (sub-mSv) CT colonography: a prospective comparison of image quality and polyp conspicuity at reduced-dose versus standard-dose imaging. *Eur Radiol* 2015; 25:2089–2102
11. Hérin E, Gardavaud F, Chiaradia M, et al. Use of model-based iterative reconstruction (MBIR) in reduced-dose CT for routine follow-up of patients with malignant lymphoma: dose savings, image quality and phantom study. *Eur Radiol* 2015; 25:2362–2370
12. Boudabbous S, Arditi D, Paulin E, Syrogiannopoulou A, Becker C, Montet X. Model-based iterative reconstruction (MBIR) for the reduction of metal artifacts on CT. *AJR* 2015; 205:380–385
13. Wang R, Sui X, Schoepf UJ, et al. Ultralow-radiation-dose chest CT: accuracy for lung densitometry and emphysema detection. *AJR* 2015; 204:743–749
14. Ebner L, Bütikofer Y, Ott D, et al. Lung nodule detection by microdose CT versus chest radiography (standard and dual-energy subtracted). *AJR* 2015; 204:727–735
15. Huber A, Landau J, Ebner L, et al. Performance of ultralow-dose CT with iterative reconstruction in lung cancer screening: limiting radiation exposure to the equivalent of conventional chest x-ray imaging. *Eur Radiol* 2016; 26:3643–3652
16. Messerli M, Kluckert T, Knitel M, et al. Ultralow dose CT for pulmonary nodule detection with chest x-ray equivalent dose: a prospective intraindividual comparative study. *Eur Radiol* 2017; 27:3290–3299
17. Hata A, Yanagawa M, Kikuchi N, Honda O, Tomiyama N. Pulmonary emphysema quantification on ultra-low-dose computed tomography using model-based iterative reconstruction with or without lung setting. *J Comput Assist Tomogr* 2018; 42:760–766
18. Hata A, Yanagawa M, Honda O, Miyata T, Tomiyama N. Ultra-low-dose chest computed tomography for interstitial lung disease using model-based iterative reconstruction with or without the lung setting. *Medicine (Baltimore)* 2019; 98:e15936
19. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6:837–845
20. Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* 2019; 20:938–947
21. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019; 394:861–867
22. Tian S, Liu A, Liu J, Liu Y, Pan J. Potential value of the PixelShine deep learning algorithm for increasing quality of 70 kVp+ASiR-V reconstruction pelvic arterial phase CT images. *Jpn J Radiol* 2019; 37:186–190
23. Aberle DR, Adams AM, Berg CD, et al.; National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365:395–409
24. Pinsky PF, Gierada DS, Black W, et al. Performance of lung-RADS in the national lung screening trial: a retrospective assessment. *Ann Intern Med* 2015; 162:485–491
25. American Association of Physicists in Medicine (AAPM) website. AAPM Task Group 23. The measurement, reporting, and management of radiation dose in CT. aapm.org/pubs/reports/RPT_96.pdf. Published January 2008. Accessed September 14, 2020
26. Hata A, Yanagawa M, Honda O, Gyobu T, Ueda K, Tomiyama N. Submillisievert CT using model-based iterative reconstruction with lung-specific setting: an initial phantom study. *Eur Radiol* 2016; 26:4457–4464
27. van Riel SJ, Jacobs C, Scholten ET, et al. Observer variability for Lung-RADS categorisation of lung cancer screening CTs: impact on patient management. *Eur Radiol* 2019; 29:924–931
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44:837–845
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174
30. American College of Radiology (ACR) website. Lung-RADS version 1.1 assessment categories. www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf. Published 2019. Accessed December 9, 2019