European Radiology (2020) 30:487–500
https://doi.org/10.1007/s00330-019-06359-6

PHYSICS

CT iterative reconstruction algorithms: a task-based image quality assessment

J. Greffier[1] · J. Frandon[1] · A. Larbi[1] · J. P. Beregi[1] · F. Pereira[1]

Received: 7 May 2019 / Revised: 21 June 2019 / Accepted: 3 July 2019 / Published online: 29 July 2019

Abstract

Purpose To assess the dose performance in terms of image quality of filtered back projection (FBP) and two generations of iterative reconstruction (IR) algorithms developed by the most common CT vendors.

Materials and methods We used four CT systems equipped with a hybrid/statistical IR (H/SIR) and a full/partial/advanced model-based IR (MBIR) algorithms. Acquisitions were performed on an ACR phantom at five dose levels. Raw data were reconstructed using a standard soft tissue kernel for FBP and one iterative level of the two IR algorithm generations. The noise power spectrum (NPS) and the task-based transfer function (TTF) were computed. A detectability index (d′) was computed to model the detection task of a large mass in the liver (large feature; 120 HU and 25-mm diameter) and a small calcification (small feature; 500 HU and 1.5-mm diameter).

Results With H/SIR, the highest values of d′ for both features were found for Siemens, then for Canon and the lowest values for Philips and GE. For the large feature, potential dose reductions with MBIR compared with H/SIR were − 35% for GE, − 62% for Philips, and − 13% for Siemens; for the small feature, corresponding reductions were − 45%, − 78%, and − 14%, respectively. With the Canon system, a potential dose reduction of − 32% was observed only for the small feature with MBIR compared with the H/SIR algorithm. For the large feature, the dose increased by 100%.

Conclusion This multivendor comparison of several versions of IR algorithms allowed to compare the different evolution within each vendor. The use of d′ is highly adapted and robust for an optimization process.

Key Points

- *The performance of four CT systems was evaluated by using imQuest software to assess noise characteristic, spatial resolution, and lesion detection.*
- *Two task functions were defined to model the detection task of a large mass in the liver and a small calcification.*
- *The advantage of task-based image quality assessment for radiologists is that it does not include only complicated metrics, but also clinically meaningful image quality.*

Keywords Multidetector computed tomography · Image enhancement · Image reconstruction

Abbreviations

CTDI	CT dose index
$CTDI_{vol}$	Volume CT dose index
ESF	Edge-spread function
FBP	Filtered back projection
H/SIR	Hybrid or statistical iterative reconstruction
IR	Iterative reconstruction
LSF	Line-spread function
MBIR	Full or advanced or partial model-based iterative reconstruction
NPS	Noise power spectrum
TTF	Task-based transfer function

✉ J. Greffier
joel.greffier@chu-nimes.fr

[1] Department of Radiology, CHU Nîmes, Medical Imaging Group Nîmes, University of Montpellier, EA 2415, Bd Prof Robert Debré, 30029 Nîmes Cedex 9, France

Introduction

Iterative reconstruction (IR) algorithms were introduced in 2008 to improve image quality and therefore the accuracy of diagnosis. Compared with filtered back projection (FBP), IR

algorithms decrease noise and also maintain image quality with reduced doses [1–5].

The most common CT vendors have produced several generations of IR algorithms. The most recent are hybrid/statistical IR (H/SIR) and full/partial/advanced model-based iterative reconstruction (MBIR) algorithms [6, 7]. H/SIR combines FBP and IR in different proportions to achieve reconstruction. MBIR uses a probabilistic method, deriving a statistical cost function by incorporating X-ray physics and computed tomography (CT) optics modeling to reduce noise and artifacts [8, 9]. Reconstruction with the full MBIR algorithm is time-consuming (e.g., Veo, GE), but a faster partial/advanced version has been developed [7].

Using IR modifies the evaluation of image quality parameters such as image noise, contrast-to-noise ratio, and modulation transfer function. Indeed, different studies have shown that the contrast-to-noise ratio was improved using IR reconstruction, although the image texture changes. Therefore, complex image quality properties of IR images require the use of adapted metrics. The noise power spectrum (NPS) is computed to evaluate noise components, notably texture, which differs between IR and FBP images. Using IR reconstruction, the NPS peak was reduced and NPS spatial frequency was shifted toward a lower frequency. As defined by Verdun et al, non-linear and non-stationary properties make spatial resolution dependent on contrast and dose [10]. To address this problem, a task-based transfer function (TTF) adapted to each clinical task is used.

However, considerations about metrics are not an important concern for radiologists, who focus on identifying/locating/characterizing abnormal images. Image quality is usually evaluated subjectively. To investigate the relationship between objective and subjective metrics without assessment by a radiologist, complex metrics such as the detectability index (d′) are required [11]. The d′ estimates the radiologist's ability to perform a clinical task. It corresponds to a figure of merit reflecting the resolution and noise properties (TTF and NPS outcomes) of IR algorithms as they relate to the ability of the system to perform a task of interest.

The goal of d′ is to find the lowest dose that produces images of sufficient quality to provide good clinical performance. Thus, d′ serves to measure the improvement in detectability at a given dose and to determine the dose reduction that ensures similar detection performance with FBP versus IR or between two IR algorithms [11].

Few studies have been published using the d′ metric to evaluate distinct IR algorithms. Samei and Richards compared three reconstruction types: FBP, H/SIR (Asir algorithm), and full MBIR (Veo algorithm) of the same manufacturer (GE Healthcare) [11]. This task-based image quality assessment performed for two detection tasks (detection of a relatively small and a relatively large feature, 1.5 and 25 mm respectively) was a robust and complete method to compare the performance of IR algorithms.

The aim of our study was to assess the dose performance in terms of image quality with two IR algorithm generations developed by the four most common CT vendors using the two clinical tasks previously defined by Samei and Richards [11].

Materials and methods

CT systems

CT systems produced by the main four manufacturers were selected: Revolution GSI (GE Healthcare), Ingenuity iCT (Philips Medical Systems), Somatom Definition AS+ (Siemens Healthineers), and Aquilion One Genesis (Canon Medical Systems). Table 1 presents the two generations of IR algorithms of each CT system assessed. The type of IR algorithms (hybrid or MBIR) was classified according to the classification of Willemink and Noël [12].

Phantom

A 20-cm-diameter ACR QA phantom (Gammex 464) placed inside a body ring (diameter of 33 cm and length of 24 cm) was used to measure IR-appropriate physical metrics such as NPS and TTF (Fig. 1). TTF was computed in module 1. This module was composed of four inserts of 25-mm diameter each that were placed into a water equivalent as background material (HU between − 7 and 7). Bone (HU between 850 and 970) and acrylic (HU between 110 and 135) inserts were used. NPS was computed in module 3. This module consists of a uniform, water-equivalent material with two very small breads of 28 mm. The thickness of both modules was 40 mm.

Acquisition and reconstruction parameters

Tube voltage was set at 100 kV. Tube currents (mA) were defined to obtain five dose levels: 0.5, 1.5, 3.0, 7.0, and 12.0 mGy. The first two levels were those used in our ultra-low-dose protocols for thoracic and abdominal-pelvic CT. The other levels were those currently employed at our institution for thoracic, abdominal-pelvic, and lumbar spine acquisitions. All acquisitions were performed with a rotation time of 0.5 s/rot and disabling of tube current modulation.

Raw data were reconstructed using a standard soft tissue reconstruction kernel combined with FBP and an intermediate iterative level of the H/SIR and MBIR algorithms. Concerning the Siemens machine, acquisitions were performed on the same CT before and after the upgrade of hardware and software components. The Ultra-Fast-Ceramic™ detector was changed to a Stellar detector, and SAFIRE was upgraded to

Table 1 Acquisition and reconstruction parameters used for each CT scan

Manufacturer		GE	Philips	Siemens	Canon
Model		Revolution GSI	iCT	Definition AS+	Aquilion One Genesis
mAs values used according to $CTDI_{vol}$	0.50 mGy	0.51 mGy (20 mAs)	0.50 mGy (12 mAs)	0.51 mGy (13 mAs)	0.50 mGy (10 mAs)
	1.50 mGy	1.54 mGy (60 mAs)	1.50 mGy (34 mAs)	1.50 mGy (38 mAs)	1.50 mGy (40 mAs)
	3.00 mGy	2.95 mGy (115 mAs)	3.00 mGy (68 mAs)	2.99 mGy (76 mAs)	3.00 mGy (80 mAs)
	7.00 mGy	6.94 mGy (270 mAs)	7.00 mGy (160 mAs)	7.01 mGy (178 mAs)	7.10 mGy (190 mAs)
	12.00 mGy	11.89 mGy (465 mAs)	12.00 mGy (274 mAs)	12.00 mGy (305 mAs)	12.00 mGy (305 mAs)
Pitch		0.984	0.984	1.000	0.813
First-generation studied	Algorithm name	Asir	iDose[4]	SAFIRE	AIDR3D
	Type of algorithm	Hybrid	Hybrid	Hybrid	Hybrid
	Level	60%	Level 4	Level 3	Standard
	Reconstruction kernel	Standard	B	I30f	FC13
	Detector	Gemstone	Elite IMR Ready	Ultra Fast Ceramic	PURE Vision
Second-generation studied	Algorithm name	Asir-V	IMR	ADMIRE	First
	Type of algorithm	Partial model-based	Full model-based	Advanced model-based	Advanced model-based
	Level	60%	Level 2	Level 3	Standard
	Reconstruction kernel	Standard	Routine	I30f	Body Sharp
	Detector	Gemstone	Elite IMR Ready	Stellar	PURE Vision
Thickness/overlapped		1.25 mm/1.25 mm	1 mm/1 mm	1 mm/1 mm	1 mm/1 mm
Collimation		64 × 0.625 mm	64 × 0.625 mm	64 × 0.6 mm	64 × 0.625 mm
Pixel size (pitch) of detector in *x* and *z* directions (mm)		1.000 × 1.000	1.408 × 1.140	1.273 × 1.095	Not available
Pixel size of axial image (mm)		0.488	0.488	0.488	0.488
Size of focus (IEC 60336)		1.6 × 1.2	1.1 × 1.2	0.9 × 1.1	1.6 × 1.5
%$CTDI_{vol}$ displayed and measured for large collimation and 32-cm reference phantom		8.7%	6.0%	9.2%	3.3%

ADMIRE advanced modeled iterative reconstruction; *AIDR3D* adaptive iterative dose reduction 3D; *Asir* adaptive statistical iterative reconstruction; *First* forward projected model-based iterative reconstruction solution; *iDose* intelligent dose; *IMR* iterative model reconstruction; *SAFIRE* sinogram-affirmed iterative reconstruction

ADMIRE. In clinical practice, SAFIRE could be used with UFC or Stellar detectors but ADMIRE only with Stellar detectors. It is important to highlight that SAFIRE and ADMIRE are not available simultaneously on the same CT. The other systems have two IR algorithms currently used in clinical practice.

Images were reconstructed with a field of view of 250 mm and slice thickness close to 1 mm (1-mm increment). Table 1 reports the acquisition and reconstruction parameters used in this study.

Dosimetry

Volume CT dose indexes ($CTDI_{vol}$), determined for a 32-cm-diameter (polymethyl methacrylate) reference phantom, were retrieved from the review report available in the CT workstation at the end of the acquisitions. The deviation between the $CTDI_{vol}$ measured during the annual mandatory control quality and displayed for all CT systems used was lower than 25% according to IEC 60601-2-44 [13].

Image quality assessment

Image quality assessments were carried out using imQuest software (12). imQuest is an image analysis tool designed to facilitate task-based image quality assessment of CT images. The tool allows one to measure resolution in terms of a TTF, noise in terms of a NPS, and detectability index for several variants of a non-prewhitening (NPW) matched filter observer model. The methods used by the software have been used in various other peer-reviewed papers [11, 14, 15] and are part of a forthcoming American Association of Physicists in Medicine task group report, TG-233 [14].

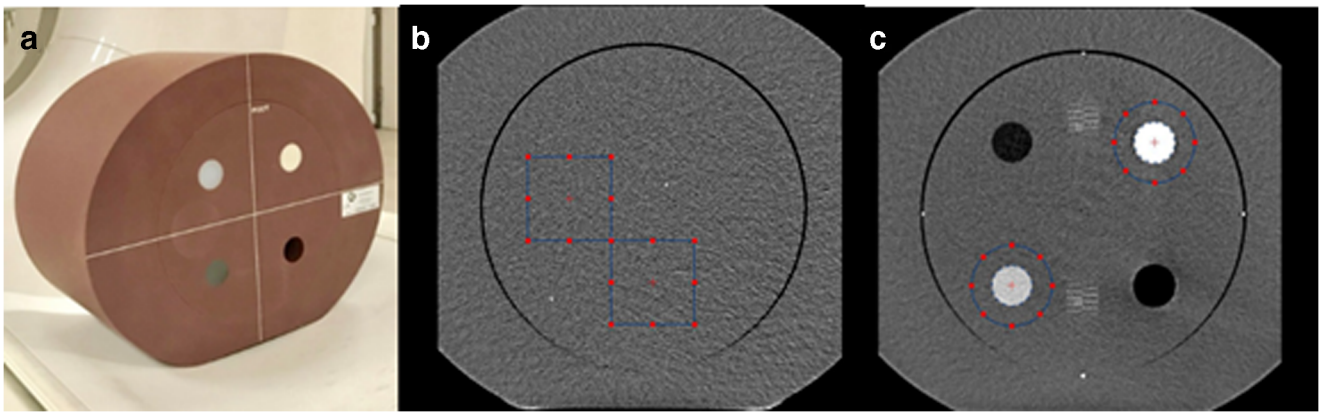

Fig. 1 **a** Phantom used in the study. **b** Regions of interest (ROIs) used for the noise power spectrum (NPS) assessment. **c** ROIs used to compute the task-based transfer function (TTF) with the bone and acrylic inserts

Noise power spectrum

NPS was computed by placing two square regions of interest (ROIs) in the uniform section (module 3) of the ACR phantom (Fig. 1b) as follows:

$$\mathrm{NPS}_{2D}\left(f_x,f_y\right)=\frac{\Delta_x\Delta_y}{L_xL_y}\frac{1}{N_{\mathrm{ROI}}}\sum_{i=1}^{N_{\mathrm{ROI}}}\left|\mathrm{FT}_{2D}\left\{\mathrm{ROI}_i\left(x,y\right)-\overline{\mathrm{ROI}_i}\right\}\right|^2$$

where Δ_x and Δ_y are pixel size in the x- and y- directions (0.488 mm), L_x and L_y the ROI lengths in the x- and y- directions, N_{ROI} the number of ROIs, FT the Fourier transform, and $\overline{\mathrm{ROI}_i}$ the background or structured noise measured from ROI(x, y) using a first-order (subtraction of a 3D linear fit) detrending technique. To improve the measurement statistics [11], the ensemble NPS was computed on a total of 40 ROIs (N_{ROI}), 128 × 128 pixels (L_x and L_y) each, within 20 consecutive axial slices.

Task transfer function

TTF was assessed using two cylindrical inserts available in module 1 of the ACR phantom [11, 16]. A circular ROI was placed around the insert, and a circular-edge technique was applied to measure the edge-spread function (ESF). The ESF was obtained by measuring the radius of each pixel from the center of each pixel of the insert. To minimize the effect of noise in the ESF, the ensemble ESF was computed by averaging 10 ESF measured across 10 consecutive axial slices [16]. The line-spread function (LSF) was also obtained by derivation of the ESF ensemble. TTF was computed from the normalized Fourier transformation of the LSF.

TTF measurements were performed using acrylic and bone inserts. The value of TTF at 50% ($\mathrm{TTF}_{50\%}$) was computed.

Detectability index (d′)

Several studies performed a task-based image quality assessment using a different model observer. These differences refer to the level of the complexity and the facility of the implementation in routine. The most common model observers are the NPW observer model [11] and the channelized Hotelling observer [7]. The first one was used in this study.

The NPW observer model uses a simple template-matching strategy to determine if a given image contains the signal of interest or not. The template it uses is just the expected signal. The NPW model does not attempt to account for noise correlations. The NPW observer model with an eye filter (NPWE) has the same strategy but also incorporates a model of the human visual system and its non-uniform response to different spatial frequencies. For such a model, it is possible to predict its detectability index if you know the noise (NPS) and resolution (TTF) properties of the images and if you define the properties of the signal to be detected (the task function, $W(u,v)$).

This observer model ($\mathrm{d'}_{\mathrm{NPWE}}$) was used to calculate d′ as follows:

$$d'^2_{\mathrm{NPWE}}=\frac{\left[\iint|W(u,v)|^2.\mathrm{TTF}(u,v)^2.E(u,v)^2dudv\right]^2}{\iint|W(u,v)|^2.\mathrm{TTF}(u,v)^2.\mathrm{NPS}(u,v)^2.E(u,v)^4dudv}$$

where u and v are spatial frequencies in the x and y directions, respectively, E is the eye filter that models the human visual system's sensitivity to different spatial frequencies [14, 17–20], and $W(u,v)$ is the task function defined as:

$$W=|F\{h_1(x,y)-h_2(x,y)\}|$$

where $h_1(x,y)$ and $h_2(x,y)$ correspond to the object present and the object absent hypothesis.

The eye filter was modeled according to the visual response function [19]. Two task functions were defined to represent

large and small features, according to the task defined by Samei and Richards [11]. The large feature was assumed to represent a circular signal with a pre-imaged contrast of 120 HU and a diameter of 25 mm compared with 500 HU and 1.5 mm for the small feature. TTF outcomes of the acrylic insert were used for the large feature and those of the bone insert for the small feature. The large feature was defined to model the detection task of a large mass in the liver and the small feature to model the detection of calcifications or the identification of high-contrast tissue boundaries. The small feature can be related to the 0.7 line pair per millimeter resolution pattern available in module 4 of the ACR phantom and the large feature to the acrylic insert of module 1.

Interpretation conditions used to obtain d′ were a zoom factor of 1.5, a viewing distance of 400 mm, and a field of view of 500 mm.

For each detection task, reconstruction kernel, and protocol, the TTF and NPS were combined with the task function to estimate d′ as a function of the volume CT dose index ($CTDI_{vol}$).

Results

Image quality assessment

Noise power spectrum

Table 2 reports the NPS peak and spatial frequency data and Fig. 2 shows the NPS curves with H/SIR and MBIR at three dose levels. The NPS peak decreased as the dose increased. The NPS peak was lower with MBIR than H/SIR with the GE (−40% ± 4%) and Philips (−71% ± 11%) systems. With Siemens, the NPS peak was higher with ADMIRE 3 versus SAFIRE 3 in the 0.5- to 3-mGy dose range (0% ± 17%), whereas the opposite was true at 7 and 12 mGy. With Canon, the NPS peak was 89% ± 17% higher with First Standard compared with AIDR 3D Standard.

NPS spatial frequency data refers to the frequency at which the maximum of the NPS is reached. The NPS spatial frequency increased as the dose increased. For all systems but Siemens, NPS spatial frequencies were lower with MBIR than with

Table 2 Values of NPS peak and NPS spatial frequency for hybrid/statistical IR and full/advanced/partial MBIR algorithms according to dose level

		$CTDI_{vol}$ (mGy)				
		0.5	1.5	3	7	12
NPS peak (HU^2 mm^2)	FBP	35712	14519	7970	3304	1862
	Asir 60%	30313	11431	5856	2392	1353
	Asir-V 60%	19447	6188	3462	1428	844
	FBP	62982	21389	8215	3303	1837
	iDose[4] 4	22646	8911	4300	1913	1104
	IMR 2	10803	2710	1089	440	221
	FBP-UFC	15803	8424	4275	1834	1030
	FBP-Stellar	14200	5246	3163	1742	1134
	SAFIRE 3	10561	4523	2312	1022	583
	ADMIRE 3	9804	3641	2078	1138	720
	FBP	57098	12082	4336	1578	992
	AIDR 3D Standard	6400	2917	1765	926	691
	First Standard	13836	5432	3034	1736	1253
NPS spatial frequency (mm^{-1})	FBP	0.205	0.283	0.299	0.299	0.315
	Asir 60%	0.173	0.189	0.189	0.205	0.205
	Asir-V 60%	0.142	0.142	0.142	0.142	0.157
	FBP	0.189	0.205	0.205	0.220	0.220
	iDose[4] 4	0.094	0.173	0.173	0.189	0.189
	IMR 2	0.063	0.063	0.063	0.079	0.079
	FBP-UFC	0.157	0.189	0.189	0.205	0.220
	FBP-Stellar	0.157	0.173	0.189	0.189	0.220
	SAFIRE 3	0.126	0.142	0.142	0.157	0.157
	ADMIRE 3	0.142	0.157	0.157	0.173	0.189
	FBP	0.173	0.189	0.189	0.220	0.236
	AIDR 3D Standard	0.081	0.126	0.126	0.173	0.189
	First Standard	0.079	0.094	0.110	0.142	0.142

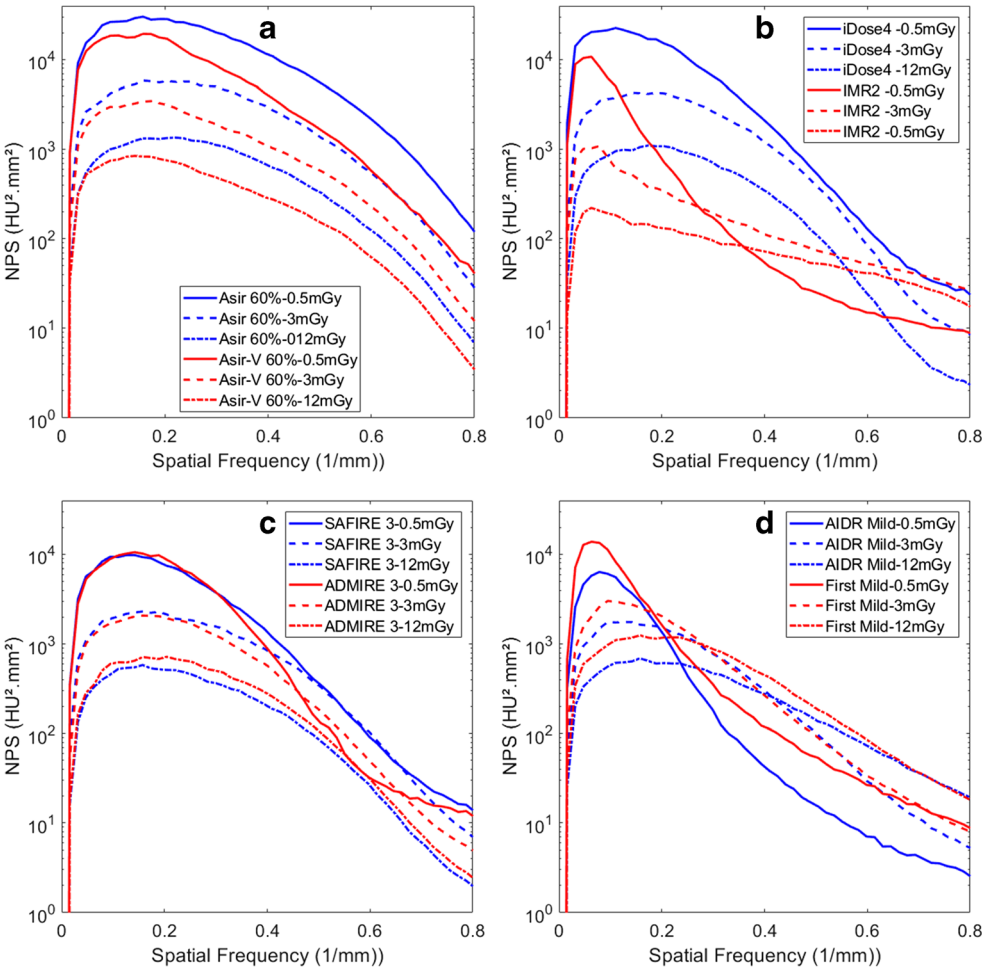

Fig. 2 NPS curves obtained at three dose levels (0.5, 3, and 12 mGy) with the H/SIR and MBIR algorithms. **a** GE **b** Philips **c** Siemens **d** Canon

H/SIR, with mean reductions of − 24% ± 5% with GE, − 55% ± 13% with Philips, and − 16% ± 10% with Canon. With Siemens, the NPS spatial frequency was 13% ± 4% higher with ADMIRE 3 versus SAFIRE 3.

Task transfer function

Figure 3 depicts the TTF curves for MBIR and H/SIR at three dose levels with the acrylic (Fig. 3a) and bone inserts (Fig. 3b). Table 3 reports the $TTF_{50\%}$ values for both inserts. TTF values tended to decline as dose decreased (Fig. 3). With GE and the acrylic insert (Table 2), $TTF_{50\%}$ was lower by − 17% ± 8% with Asir-V 60% versus Asir 60%. With GE and the bone insert, $TTF_{50\%}$ was higher with Asir-V 60% than with Asir 60% at 0.5, 1.5, and 3 mGy, whereas the opposite occurred with 7 and 12 mGy. With Philips, $TTF_{50\%}$ was lower with IMR 2 than iDose^4 4 for the low-dose levels (0.5 to 3 mGy with the acrylic insert and 0.5 to 1.5 mGy with the bone insert), but the opposite occurred at higher doses. Findings were similar for the acrylic insert with the Siemens and Canon systems. $TTF_{50\%}$ was lower with H/SIR versus MBIR at 0.5 and 1.5 mGy with Siemens and at 0.5 to 3 mGy with Canon. With the bone insert, $TTF_{50\%}$ was higher with MBIR than H/SIR at all dose levels, showing mean increases of 7% ± 4% with Siemens and 76% ± 15% with Canon.

Detectability index (d′)

Comparison of CT systems Figure 4 depicts d′ values obtained with FBP, H/SIR, and MBIR. With FBP, d′ values were higher with Siemens than with the other three systems. The use of Stellar detectors improved d′ values at dose levels below 7 mGy compared with UFC detectors (14% ± 5% for large feature and 5% ± 6% for small feature), whereas at 7 and 12 mGy, d′ was 1% and 8% higher using UFC than Stellar detectors, respectively. The d′ values were on the same order of magnitude for the GE and Canon systems. Compared with both previous systems, d′ values were higher with the Philips system (15% ± 16% for large feature and 26% ± 14% for small feature). For small feature detection with H/SIR, Siemens provided the highest d′ values, followed by Canon and then GE and Philips. For the large feature, Canon produced the highest d′ values at

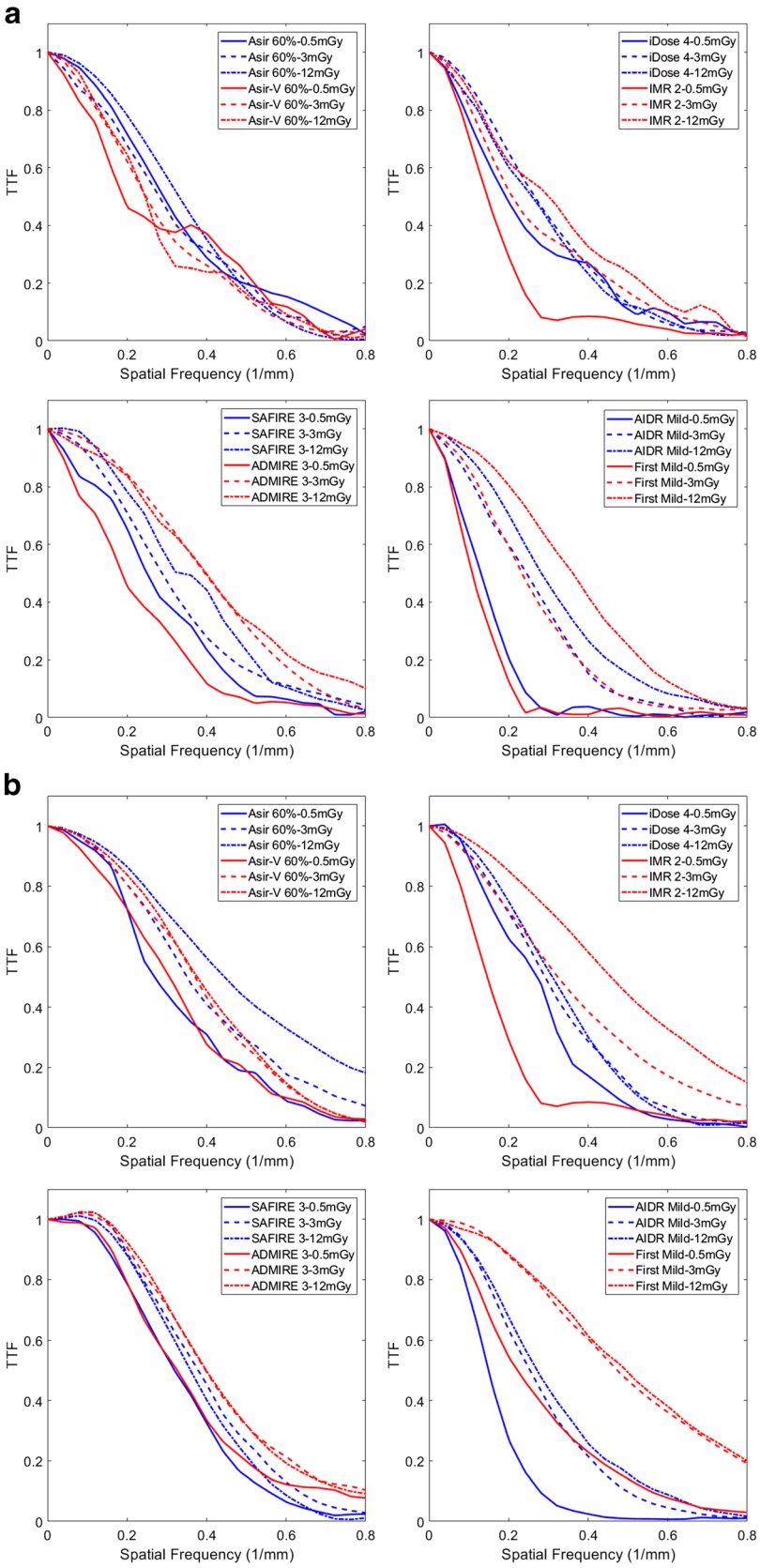

Fig. 3 TTF curves obtained at three dose levels (0.5, 3, and 12 mGy) with the H/SIR and MBIR algorithms with the acrylic (**a**) and bone (**b**) inserts. top left, GE; top right, Philips; bottom left, Siemens; bottom right, Canon

Table 3 Values of $TTF_{50\%}$ with the acrylic and bone inserts and with the hybrid/statistical IR and full/advanced/partial MBIR algorithms according to dose level

		$CTDI_{vol}$ (mGy)				
		0.5	1.5	3	7	12
$TTF_{50\%}$ acrylic	FBP	0.479	0.501	0.475	0.441	0.496
	Asir 60%	0.271	0.278	0.281	0.320	0.330
	Asir-V 60%	0.189	0.235	0.247	0.284	0.285
	FBP	0.254	0.281	0.323	0.288	0.301
	iDose^4 4	0.192	0.194	0.266	0.252	0.259
	IMR 2	0.135	0.146	0.209	0.262	0.300
	FBP-UFC	0.305	0.359	0.428	0.430	0.432
	FBP-Stellar	0.280	0.327	0.394	0.407	0.396
	SAFIRE 3	0.250	0.264	0.288	0.344	0.335
	ADMIRE 3	0.187	0.255	0.396	0.401	0.398
	FBP	0.314	0.357	0.318	0.381	0.322
	AIDR 3D Standard	0.130	0.236	0.245	0.278	0.282
	First Standard	0.109	0.217	0.234	0.347	0.362
$TTF_{50\%}$ bone	FBP	0.378	0.421	0.412	0.433	0.410
	Asir 60%	0.269	0.331	0.351	0.430	0.445
	Asir-V 60%	0.306	0.368	0.371	0.381	0.376
	FBP	0.296	0.298	0.299	0.302	0.307
	iDose^4 4	0.271	0.276	0.292	0.304	0.310
	IMR 2	0.191	0.263	0.324	0.425	0.457
	FBP-UFC	0.285	0.344	0.343	0.351	0.349
	FBP-Stellar	0.262	0.300	0.317	0.327	0.319
	SAFIRE 3	0.319	0.348	0.379	0.372	0.359
	ADMIRE 3	0.326	0.379	0.397	0.403	0.399
	FBP	0.368	0.327	0.341	0.328	0.296
	AIDR 3D Standard	0.150	0.221	0.254	0.278	0.272
	First Standard	0.223	0.399	0.473	0.501	0.495

0.5 to 3 mGy and Siemens at 7 and 12 mGy. With GE, the d′ values were lower than with Canon and Siemens but higher than with Philips (9% ± 3%). With MBIR, d′ values for both features were higher with Philips than with the other systems (large feature, 54% ± 5% vs Asir-V; 14% ± 11% vs ADMIRE, 51% ± 26% vs First and 98% ± 16%, 30% ± 12% and 68% ± 45% for the small feature, respectively). Siemens produced higher d′ values compared with GE and Canon for both features. Finally, with Canon, d′ values were higher than with GE from 0.5 to 3 mGy but lower at 12 mGy.

Comparison of IR algorithms with each CT system Figure 5 reports d′ for the small (Fig. 5a) and large (Fig. 5b) features. With GE and Philips, d′ values were higher with MBIR compared with H/SIR. The difference was greater for the small feature for both systems and with the Philips MBIR algorithm. With Siemens, d′ was higher with ADMIRE 3 than with SAFIRE 3 at low-dose levels, whereas the opposite occurred at higher dose levels. The inversion point was close to 5 mGy for the large feature and 8 mGy for the small feature. With Canon, d′ values for the large feature were lower with First Standard compared with AIDR 3D Standard, whereas the opposite was noted for the small feature. These outcomes were associated with the TTF and NPS results, chiefly the spatial frequency values.

Potential increase in d′ and potential dose reduction To evaluate the potential increase in d′ with MBIR versus H/SIR, we compared d′ values at a nominal $CTDI_{vol}$ level of 3 mGy. Table 4 shows that MBIR improved the detectability of both features with all systems except Canon for the large feature (− 22%). The d′ increase was greater for the small feature and with Philips (69% for the large feature and 165% for the small feature) compared with the other three systems.

To evaluate potential dose reduction, we compared the $CTDI_{vol}$ obtained when MBIR was used to achieve the same d′ value as with H/SIR at 3 mGy. Using MBIR reduced the

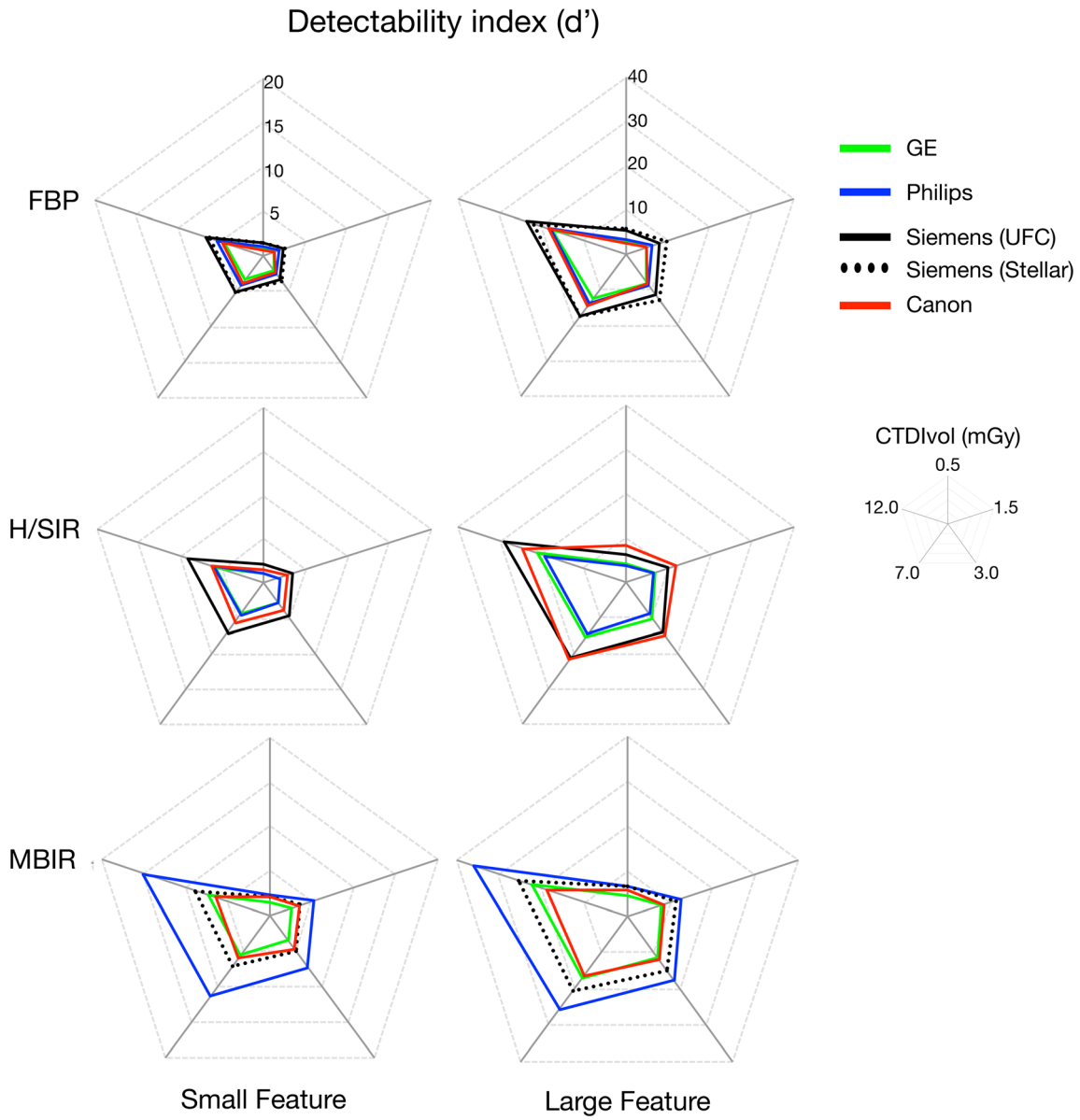

Fig. 4 Comparisons of detectability index (d′) values obtained for each reconstruction type at each dose level with the four CT systems

$CTDI_{vol}$ for both features with all systems except Canon for the large feature (100%). This dose reduction was greater for the small feature and with Philips (− 62% for the large feature and − 78% for the small feature).

Subjective assessment of image quality

Figure 6 depicts the image quality obtained with the acrylic insert at dose levels and with all H/SIR and MBIR algorithms. For all systems, image noise increased with the dose reduction. Compared with Asir, Asir-V reduces the image noise, which allowed improvement in spatial resolution (visual border detection) and contrast. The detectability of the insert was difficult at 0.5 mGy for both IR algorithms. Using Asir-V, images are more smoothed at 7 and 12 mGy.

Regarding the Philips system, IMR strongly reduced image noise, but the images were very smoothed. For all dose levels, image quality obtained with IMR was adapted to detect the acrylic insert. The deterioration of contrast and spatial resolution at 0.5 and 1.5 mGy limited the insert detection with iDose[4].

Concerning the Siemens system, noise, contrast, and spatial resolution were similar for both IR algorithms.

Regarding the Canon system, similar image quality was found for AIDR 3D and First for dose levels higher than 3 mGy. In addition, images were very smoothed for 0.5 mGy for both IR algorithms and 1.5 mGy for First. In both

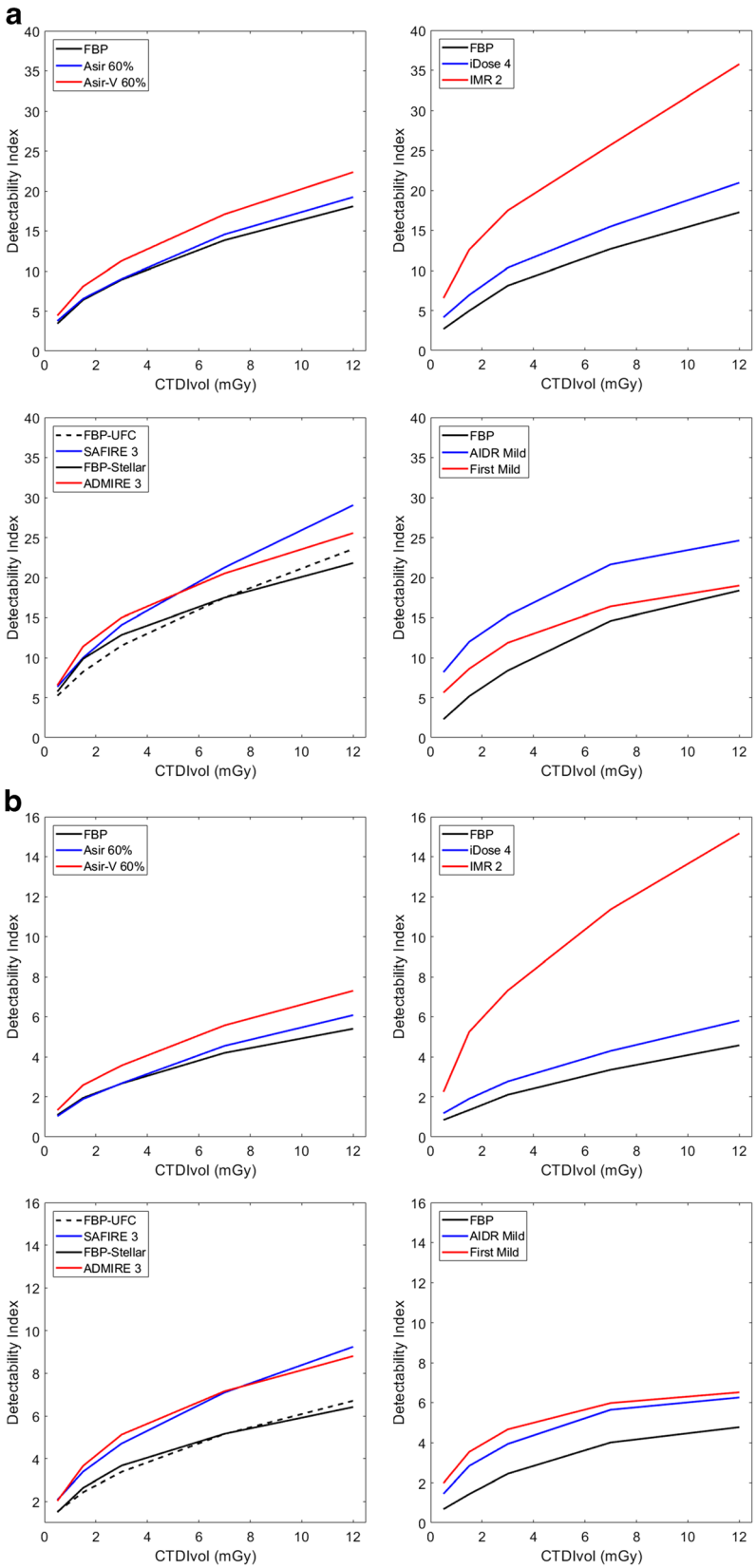

Fig. 5 Detectability index (d′) as a function of dose with the H/SIR and MBIR algorithms for detection of the small feature (1.5 mm in diameter, 500 HU contrast) (**a**) and large feature (25-mm diameter, 120 HU contrast) (**b**)

Table 4 Image quality improvement and potential dose reduction with full/advanced/partial MBIR compared with hybrid/statistical IR (H/SIR). The "increase in d′" data correspond to a nominal $CTDI_{vol}$ of 3 mGy and the "dose reduction" data to the d′ value obtained with hybrid/statistical IR at 3 mGy

	Large feature		Small feature	
	Increase in d′ (%)	Dose reduction (%)	Increase in d′ (%)	Dose reduction (%)
Asir 60% to Asir-V 60%	25%	− 35%	33%	− 45%
iDose^4 4 to IMR2	69%	− 62%	165%	− 78%
SAFIRE 3 to ADMIRE 3	7%	− 13%	9%	− 14%
AIDR 3D Standard to First Standard	− 22%	100%	19%	− 32%

cases, the spatial resolution and contrast limit the insert detection.

Noise magnitude was greater in H/SIR than MBIR images. This parameter was more obvious visually with Philips and Canon than with GE and Siemens. Similar patterns were found for noise texture.

Discussion

Clinical and experimental studies have established that IR is effective in reducing CT radiation doses [1–5]. However, when evaluating images obtained using IR algorithms, specific criteria must be taken into account (e.g., non-linear and non-stationary properties of IRs). The software imQuest can be used to compute d′ and quantify the detectability of a clinical feature at a given dose [11, 14]. Several generations of IR algorithms have been developed by the most common CT vendors and are used in clinical practice [6, 7]. However, few studies have assessed their influence on dose performance in terms of image quality. We sought to fill this gap in knowledge by studying two IR algorithm generations developed by four CT manufacturers.

NPS reflected the differences in the design strategies used for two generations of algorithms and by the four manufacturers. With the GE system, Asir-V moderately diminished noise magnitude and also reduced spatial frequency compared with Asir. Decreasing spatial frequency changes noise texture by increasing image smoothing, thereby affecting the relative d′ values provided by the IR algorithms at a given noise level and for a given task [11]. The pattern was similar with Philips, although decreases in both parameters were more marked with IMR. Similar differences were also observed with H/SIR faced to FBP. For Siemens and Canon, whose NPS peak values were already low with H/SIR, a different strategy was applied. For instance, noise magnitude with Siemens was not reduced and was similar overall to that with ADMIRE and SAFIRE. However, there was an increase in NPS spatial frequency, which diminished the impact of ADMIRE on noise texture and reduced image smoothing versus SAFIRE. With Canon, noise magnitude was higher with First versus AIDR 3D, whereas spatial frequency was lower. Similar NPS outcomes were found in the literature but were not always obtained under identical conditions (e.g., phantom type and size, iterative level) [11, 21–25].

TTF outcomes in this study were consistent with earlier data [11, 26]. A distinction with FBP is that spatial resolution with IR algorithms lays on contrast and dose. This dependency was more evident for the acrylic insert with full/partial/advanced MBIR algorithms than with H/SIR (but not for GE). At doses below 3 mGy and low contrast (acrylic insert), the noise level was high and the rendition of spatial details consequently reduced. Under these conditions, TTF values were lower with MBIR compared with H/SIR. The opposite pattern occurred at the higher dose. With high contrast (bone insert), TTF values were higher with MBIR compared with H/SIR for Siemens or Canon systems or similar for the others [11].

The d′ values were better with IR algorithms compared with FBP. However, the outcomes of our study show that for the GE system, d′ improves only by few percentage points when using ASIR (5% ± 3% for the large feature and 3% ± 8% for the small feature). Differences in d′ values between MBIR and H/SIR algorithms varied across manufacturers. With GE and Philips, d′ values increased moderately with Asir-V (10% ± 4% for the large feature and 27% ± 8% for the small feature) and markedly with IMR (84% ± 9% for the large feature and 154% ± 24% for the small feature). These results were related to the NPS for these two manufacturers, especially the moderate and strong reduction of noise magnitude and noise texture. Similar results have been reported for Asir compared with FBP and for the MBIR algorithm Veo [11]. With Siemens, the relatively small variations in NPS and TTF resulted in smaller d′ differences between SAFIRE and ADMIRE (2% ± 10% for the large feature and 2% ± 6% for the small feature). However, d′ values were higher with ADMIRE compared with SAFIRE at low-dose levels, and vice versa at high-dose levels. It is important to note that for Siemens, d′ values started at a better initial value in FBP compared with other manufacturers. With Canon, d′ values were higher with First versus AIDR 3D for the small feature, whereas the opposite occurred for the large feature. These

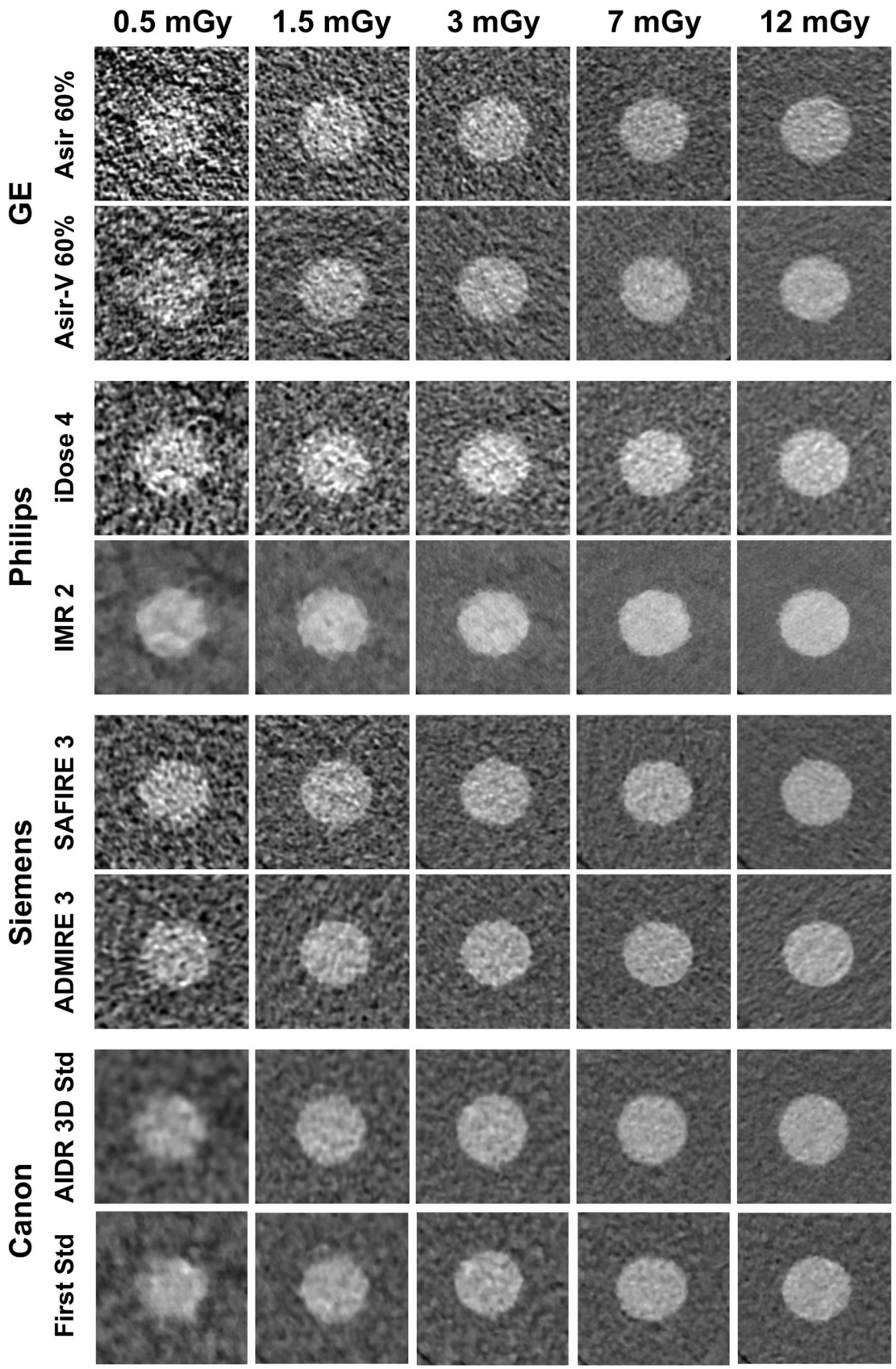

Fig. 6 A 4 × 4 cm^2 region of interest centered on the acrylic insert with the H/SIR and MBIR algorithms and the four CT systems, as a function of dose level

differences were ascribable to differences in TTF results between the two IR algorithms. With the bone insert, spatial resolution was considerably better with First compared with AIDR 3D but spatial resolution was similar between both algorithms with the acrylic insert.

Finally, the largest potential dose reduction, computed at a nominal $CTDI_{vol}$ level of 3 mGy, while keeping d′ unchanged, and the largest d′ increase while keeping the dose unchanged, was seen with IMR compared with iDose[4]. With GE and Siemens, the potential dose reduction was moderate or small with MBIR versus H/SIR. With Canon, a potential dose reduction was achieved only with First versus AIDR 3D for the small feature.

The values of d′ can be used to rank reconstruction techniques. Siemens produced the highest d′ with FBP and H/SIR for both (the large and the small features) and the second highest d′ values with MBIR, after Philips system. In addition, Canon and Siemens systems produced similar d′ values with H/SIR and the large feature. Noticeable differences of d′ values among manufacturers were found for MBIR algorithms especially for the higher dose. d′ values ranged from 1.3 to 2.2 for small feature and from 4.4 to 6.6 for large feature at 0.5 mGy and from 6.5 to 15.2 and from 19.0 to 35.8 at 12 mGy, respectively. However, this comparison of d′ values between manufacturers is limited. As shown in Table 1, the CT scans used have different intrinsic parameters (size of focus, pixel size of detector) and the parameters of the protocols are slightly different. These differences can influence the spatial resolution of the system and the amount of image noise and thus the values of d′.

This recently introduced task-based image quality assessment method helps medical physicists to determine the best compromise between dose and image quality when seeking to perform a specific clinical task (to detect a circular low- and high-contrast pattern of different sizes in this study). However, it must be combined with a radiologist evaluation of image quality in patients, notably to assess variations in noise texture and spatial resolution using lower dose levels and MBIR algorithms. The acrylic insert images, which can be likened to the large feature (Fig. 6), show that MBIR smoothed the image despite the increase in d′. In particular for GE and Philips, this modification in image texture may be disturbing for the radiologist seeking to establish a diagnosis.

Recently, at least two manufacturers have proposed artificial intelligence algorithms for the reconstruction process [12, 27]. These algorithms are based on deep learning (DL) methods, which may improve image quality as well as reduce radiation dose. For instance, General Electric proposed a DL algorithm named TrueFidelity that subtracts image noise from an image database of patients reconstructed with FBP. Similarly, Canon Medical system proposed the DL reconstruction method AiCE (Advanced Intelligent Clear-IQ Engine), which used a similar strategy although the image database of patients was reconstructed with First. A first clinical study showed that AiCE improved the quality of abdominal images of ultra-high-resolution CT [27]. To date, no studies have compared the impact of these DL methods on dose reduction and image quality compared with IR algorithms. In addition, the task-based image quality assessment evaluated in the present study might also be a subject for further comparison.

This study has several limitations. Raw data were reconstructed using a single-kernel and a single-iterative level. Other combinations of parameters may have produced different outcomes. We evaluated only two task functions, which were not representative of the range of tasks that must be performed in clinical practice. Finally, we used an image quality phantom and the results might have been different had we imaged patients.

Conclusion

The performance of each of the four imaging systems was accurately evaluated by using imQuest software to determine NPS and TTF. The d′ value was used to quantify the potential dose reduction with the H/SIR and MBIR algorithms for two clinical tasks. The use of IR algorithms such as MBIR and H/SIR in an optimization process is effective compared with FBP. In addition, the impact of MBIR compared with H/SIR depends on the manufacturers. Faced with H/SIR, MBIR produced small, moderate, and marked potential dose reductions with Siemens, GE, and Philips, respectively. With Canon, a potential dose reduction was possible only with the small feature using MBIR. This task-based image quality assessment method helps medical physicists to identify the lowest radiation dose that can be used to perform a clinical task successfully. However, it should be validated in patients by radiologists, notably to assess image texture and spatial resolution after dose optimization.

Acknowledgments We are deeply grateful to Dr. J. Solomon for support regarding the use of imQuest software. We would like to thank Pr H. Rousseau and Dr. J.M. Teissier for giving us permission to use their measurement results.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Jean Paul Beregi.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

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

Informed consent Written informed consent was not required for this study because it's a phantom study.

Ethical approval Institutional Review Board approval was not required because it's a phantom study.

Methodology
• experimental
• multicenter study

References

1. 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
2. Larbi A, Orliac C, Frandon J et al (2018) Detection and characterization of focal liver lesions with ultra-low dose computed tomography in neoplastic patients. Diagn Interv Imaging 99:311–320
3. Macri F, Greffier J, Pereira F et al (2016) Value of ultra-low-dose chest CT with iterative reconstruction for selected emergency room patients with acute dyspnea. Eur J Radiol 85:1637–1644
4. Yamada Y, Jinzaki M, Hosokawa T et al (2012) Dose reduction in chest CT: comparison of the adaptive iterative dose reduction 3D, adaptive iterative dose reduction, and filtered back projection reconstruction techniques. Eur J Radiol 81:4185–4195
5. Yan C, Xu J, Liang C et al (2018) Radiation dose reduction by using CT with iterative model reconstruction in patients with pulmonary invasive fungal infection. Radiology. 288:285–292
6. Patino M, Fuentes JM, Hayano K, Kambadakone AR, Uyeda JW, Sahani DV (2015) A quantitative comparison of noise reduction across five commercial (hybrid and model-based) iterative reconstruction techniques: an anthropomorphic phantom study. AJR Am J Roentgenol 204:W176–W183
7. Viry A, Aberle C, Racine D et al (2018) Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: a quantitative task-based phantom study. Phys Med 48:111–118
8. McCollough CH, Chen GH, Kalender W et al (2012) Achieving routine submillisievert CT scanning: report from the summit on management of radiation dose in CT. Radiology 264:567–580
9. Thibault JB, Sauer KD, Bouman CA, Hsieh J (2007) A three-dimensional statistical approach to improved image quality for multislice helical CT. Med Phys 34:4526–4544
10. Verdun FR, Racine D, Ott JG et al (2015) Image quality in CT: from physical measurements to model observers. Phys Med 31:823–843
11. 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
12. Willemink MJ, Noël PB (2019) The evolution of image reconstruction for CT-from filtered back projection to articial intelligence. Eur Radiol 29:2185–2195
13. IEC. 60601–2-44 (2002) Medical electrical equipment –. Part 2–44: Particular requirements for the safety of. X-ray equipment for computed tomography. Edition 2.1
14. Solomon J, Zhang Y, Wilson J, Samei E (2018) An automated software tool for task-based image quality assessment and matching in clinical CT using the TG-233 Framework, Med Phys 45:E134 - E134
15. 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
16. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E (2012) Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. Med Phys 39:4115–4122
17. Burgess AE, Li X, Abbey CK (1997) Visual signal detectability with two noise components: anomalous masking effects. J Opt Soc Am A Opt Image Sci Vis 14:2420–2442
18. Burgess AE, Wagner RF, Jennings RJ, Barlow HB (1981) Efficiency of human visual signal discrimination. Science 214: 93–94
19. Eckstein M, Bartroff J, Abbey C, Whiting J, Bochud F (2003) Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks. Opt Express 11:460–475
20. Ishida M, Doi K, Loo LN, Metz CE, Lehr JL (1984) Digital image processing: effect on detectability of simulated low-contrast radiographic patterns. Radiology 150:569–575
21. Aurumskjöld ML, Ydström K, Tingberg A, Söderberg M (2017) Improvements to image quality using hybrid and model-based iterative reconstructions: a phantom study. Acta Radiol 58:53–61
22. Euler A, Solomon J, Marin D, Nelson RC, Samei E (2018) A third-generation adaptive statistical iterative reconstruction technique: phantom study of image noise, spatial resolution, lesion detectability, and dose reduction potential. AJR Am J Roentgenol 210:1301–1308
23. Greffier J, Macri F, Larbi A et al (2016) Dose reduction with iterative reconstruction in multi-detector CT: what is the impact on deformation of circular structures in phantom study? Diagn Interv Imaging 97:187–196
24. Paruccini N, Villa R, Pasquali C, Spadavecchia C, Baglivi A, Crespi A (2017) Evaluation of a commercial model based iterative reconstruction algorithm in computed tomography. Phys Med 41: 58–70
25. Solomon J, Mileto A, Ramirez-Giraldo JC, Samei E (2015) Diagnostic performance of an advanced modeled iterative reconstruction algorithm for low-contrast detectability with a third-generation dual-source multidetector CT scanner: potential for radiation dose reduction in a multireader study. Radiology 275:735–745
26. Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR (2014) Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. Phys Med Biol 59:4047–4064
27. Akagi M, Nakamura Y, Higaki T et al (2019) Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. Eur Radiol. https://doi.org/10.1007/s00330-019-06170-3