COMPUTED TOMOGRAPHY

Evaluation of the impact of a novel denoising algorithm on image quality in dual-energy abdominal CT of obese patients

Fides R. Schwartz[1] · Darin P. Clark[2] · Francesca Rigiroli[1] · Kevin Kalisz[1] · Benjamin Wildman-Tobriner[1] · Sarah Thomas[1] · Joshua Wilson[3] · Cristian T. Badea[2] · Daniele Marin[1]

Abstract

Objectives Evaluate a novel algorithm for noise reduction in obese patients using dual-source dual-energy (DE) CT imaging.

Methods Seventy-nine patients with contrast-enhanced abdominal imaging (54 women; age: 58 ± 14 years; BMI: 39 ± 5 kg/m^2, range: 35–62 kg/m^2) from seven DECT (SOMATOM Flash or Force) were retrospectively included (01/2019–12/2020). Image domain data were reconstructed with the standard clinical algorithm (ADMIRE/SAFIRE 2), and denoised with a comparison (ME-NLM) and a test algorithm (rank-sparse kernel regression). Contrast-to-noise ratio (CNR) was calculated. Four blinded readers evaluated the same original and denoised images (0 (worst)–100 (best)) in randomized order for perceived image noise, quality, and their comfort making a diagnosis from a table of 80 options. Comparisons between algorithms were performed using paired t-tests and mixed-effects linear modeling.

Results Average CNR was 5.0 ± 1.9 (original), 31.1 ± 10.3 (comparison; $p < 0.001$), and 8.9 ± 2.9 (test; $p < 0.001$). Readers were in good to moderate agreement over perceived image noise (ICC: 0.83), image quality (ICC: 0.71), and diagnostic comfort (ICC: 0.6). Diagnostic accuracy was low across algorithms (accuracy: 66, 63, and 67% (original, comparison, test)). The noise received a mean score of 54, 84, and 66 ($p < 0.05$); image quality 59, 61, and 65; and the diagnostic comfort 63, 68, and 68, respectively. Quality and comfort scores were not statistically significantly different between algorithms.

Conclusions The test algorithm produces quantitatively higher image quality than current standard and existing denoising algorithms in obese patients imaged with DECT and readers show a preference for it.

Clinical relevance statement Accurate diagnosis on CT imaging of obese patients is challenging and denoising algorithms can increase the diagnostic comfort and quantitative image quality. This could lead to better clinical reads.

Key Points

- *Improving image quality in DECT imaging of obese patients is important for accurate and confident clinical reads, which may be aided by novel denoising algorithms using image domain data.*
- *Accurate diagnosis on CT imaging of obese patients is especially challenging and denoising algorithms can increase quantitative and qualitative image quality.*
- *Image domain algorithms can generalize well and can be implemented at other institutions.*

Keywords Multidetector computed tomography · Tomography, X-ray computed · Medical image processing · Obesity · Image quality enhancement

Fides R. Schwartz and Darin P. Clark share co-first authorship.

✉ Fides R. Schwartz
   fides.schwartz@duke.edu

1   Department of Radiology, Duke University Health System, 2301 Erwin Road, Box 3808, Durham, NC 27110, USA

2   Quantitative Imaging and Analysis Lab, Department of Radiology, Duke University, Durham, NC, USA

3   Clinical Imaging Physics Group, Durham, NC, USA

Abbreviations

BMI	Body mass index
CIPG	Clinical Imaging Physics Group
CNR	Contrast-to-noise ratio
DECT	Dual-energy CT
ICC	Intraclass correlation
kVp	Kilovolt peak
MAD	Median absolute deviation
ME-NLM	Multi-Energy Non-Local Means

Springer

MTF	Modulation transfer function
NPS	Noise power spectrum
PACS	Picture archiving and communication system
ROI	Region of interest
RSKR	Rank sparse kernel regression
SD	Standard deviation
sec	Second

Introduction

Imaging of large body habitus patients is challenging due to high image noise and can lead to delays in patients' diagnosis and treatment. Many practices do not perform dual-energy (DE) CT on obese patients (e.g., excluding patients weighing over 260 pounds (> 118 kg)) due to the additional photon starvation caused by the large body circumference [1].

Dose constraints on DECT scans of obese patients result in high levels of noise, due to photon-starvation effects, negatively impacting diagnostic performance in abdominal radiology [2, 3].

One potential solution to mitigate against noise in larger body habitus populations is by using denoising algorithms. However, one concern regarding the implementation of non-vendor-based denoising algorithms in clinical practice is the fact that the original projection data is needed for the algorithms to be applied. Since these are very large data files, they do not routinely get stored on the scanners themselves or in PACS and are deleted within approximately a week of acquisition at most institutions [4, 5]. Thus, denoising using projection data, both with vendor-based iterative reconstruction algorithms and custom algorithms, is often limited by data availability. In addition, using iterative reconstruction, there are limitations in spatial resolution for low-contrast structures and low-contrast detectability declines when radiation dose reductions exceed 25% [6]. By contrast, the image domain data from CT scans is usually archived in a PACS system for a relatively long term of several years, so denoising can be applied at any time after image data was acquired and reconstructed. This data can be heterogeneous due to different frequency filters, collimation, slice thicknesses, and convolution kernels, but this can be addressed through specialized algorithms.

A denoising method developed at Duke Radiology for the review process is based on rank-sparse kernel regression (RSKR) and applies locally adaptive noise estimation [7]. Other methods apply adaptive non-local means (NLM) filtering by adapting the strength of regularization using a noise variance map generated by propagating noise through the reconstruction model to denoise images. This can lead to a highly smoothed image impression but has been shown to improve image quality in animal and cadaver studies [8, 9].

The purpose of this study was to evaluate a novel algorithm for noise reduction in obese patients for dual-source dual-energy (DE) CT abdominal imaging. We hypothesized that multi-channel denoising methods (RSKR, multi-energy (ME)-NLM) could reduce image noise and improve both objective image quality metrics like contrast-to-noise ratio (CNR) and subjective metrics like reader satisfaction.

Materials and methods

This retrospective study was Health Insurance Portability and Accountability Act compliant and approved, with informed consent waived by the institutional review board.
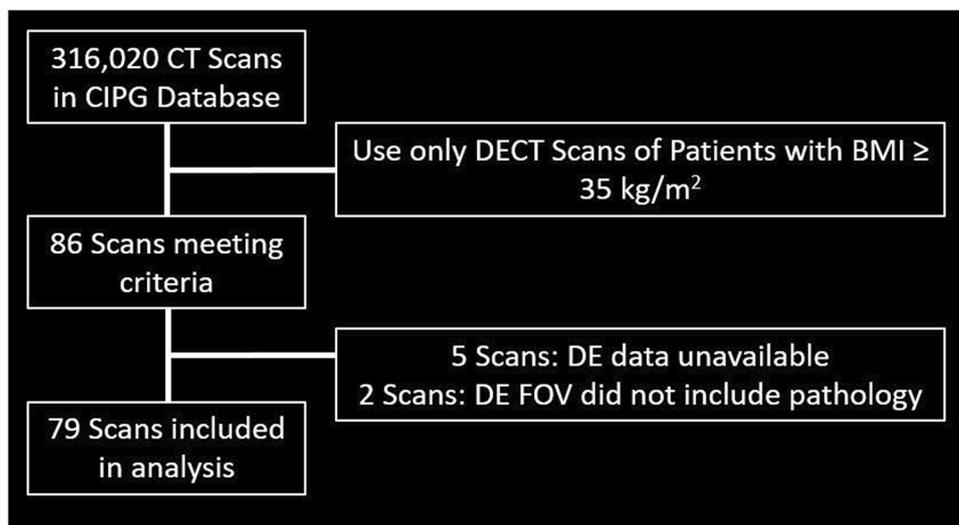
Study population

All patients with a BMI of 35 kg/m^2 or higher, found during a query of the Duke Clinical Imaging Physics Group (CIPG) Database, with examinations that included a dual-energy acquisition of the abdomen on a dual-source scanner (SOMATOM Flash or Force, Siemens Healthineers) between January 1st, 2019, and December 31st, 2020, were consecutively included in the study population (Fig. 1). Seventy-nine patients with DECT imaging from seven different scanners (54 women; age: 58 ± 14 years; BMI: 39 ± 5 kg/m^2, range: 35 to 62 kg/m^2) were included (Table 1). The scan protocols and patient radiation doses varied depending on the scanner type and location within the hospital system (Table 2).

DECT acquisition parameters

Dual-source, dual-energy CT of the abdomen was acquired on five SOMATOM Flash CT scanners with tube A set to 100 kVp and tube B at 140 kVp plus tin filtration and on two SOMATOM Force CT scanners with tube A set to 100 kVp and tube B 150 kVp plus tin filtration with a helical pitch of 0.8. Automated exposure control was used: CARE Dose with 190 quality reference mAs and CARE kV with 120 reference kV and 100–140 or 150 kV range depending on scanner type. Tube A had a field of view of 50 cm and tube B of 33 or 35 cm in the Flash and the Force scanners. All patients received 150 mL of iodinated contrast (Isovue 300, Bracco Diagnostics) at an injection rate of 4 mL/sec. Images were the acquired using bolus tracking for portal-venous phase imaging. Monitoring was performed in the descending aorta and arterial phase images were triggered at 150 HU. No further scan delay was added. Portal-venous phase was acquired 80 s after the initial trigger.

Fig. 1 The patient selection process for inclusion in our study

DECT reconstruction settings

Images were reconstructed at 1.5 mm or 0.6 mm, initially, and displayed for readers using a 1.5-mm slice and 1-mm overlap, the same as in our clinical workflow. Additional reconstruction parameters were as follows: Br40 kernel; ADMIRE/SAFIRE 2; 512×512 matrix size with associated in-plane pixel size of 0.67–0.98 mm; collimation of 19.2 or 38.4 mm at isocenter. The spectral post-processing for denoising was performed with the RSKR algorithm (test) as well as with a comparison algorithm from the literature (ME-NLM) [9] using data from tube A and tube B jointly. All images were displayed for readers as blended images with a 50/50 ratio from tube A and tube B.

The hyperparameters for denoising were chosen for RSKR via a preliminary reader preference evaluation, which was based on two radiologists trained at two different institutions in the US and Europe (filtering strength, $h_0 = 1.2$; kernel size: 13^3 voxels; filter scaling, $\gamma = 0.35$; Fig. 2). This additional step was taken to ensure that changes in NPS do not affect perceived image texture [10]. For the ME-NLM, hyperparameters were taken from the reference paper (search window, $W_\Omega = 11^3$ voxels; patch size, $W_p = 3^3$ voxels; filtering strength, $h = 1.2$). The same noise map, derived from the local median absolute deviation (MAD) of image gradients, was used for both denoising algorithms.

Quantitative image quality analysis

Quantitative metrics were computed and compared between the original, comparison, and test algorithms. Contrast-to-noise ratio (CNR) was calculated for paraspinal skeletal muscle and ventral subcutaneous fat at the same level, chosen because it was present in all images, via the formula: CNR $= |\text{attenuation}_{\text{muscle}} - \text{attenuation}_{\text{fat}}| / \sqrt{\text{SD}_{\text{muscle}}^2 + \text{SD}_{\text{fat}}^2}$. Measurements were made by DPC, with measurements defined in fat and muscle using intensity cut-off values of -200 to -50 HU for fat and -20 to 120 for soft tissue, including all voxels in the specified intensity ranges. The included voxels were applied in identical fashion to all three algorithms.

Modulation transfer function (MTF) measurements to assess spatial resolution were taken at the skin-air boundary,

Table 1 Demographic description of patient cohort

Demographic parameter	Mean $\pm$ SD
Age (years)	58 ± 14
Height (cm)	165 ± 12
Weight (kg)	108 ± 21
BMI (kg/m^2)	39 ± 5
Women (number)	54
Men (number)	25

Table 2 The distribution of scanners within the hospital, the protocol distribution on each scanner type, and the radiation dose by scanner type. $*p < 0.05$

	Somatom Flash	Somatom Force
Number of scanners	5	2
Located in emergency department	1	0
Located in Cancer Center	2	1
Located in non-specialized settings	2	1
Abdomen-pelvis protocols	28	0
GI protocols	2	1
GU protocols	6	8
Vascular protocols	15	19
Mean CTDI$_{\text{vol}}$ (mGy)	15.2 ± 7.3	$18.9 \pm 6.7*$

Singular Vectors

(1) Singular Value Decomposition

Input Data: kVp 1, kVp 2

(3) Data Fidelity

(2) Denoising

Joint Bilateral Filtration

high contrast   low contrast   noisy

Weight 1 ... 0

Output Data: kVp 1, kVp 2

(3) Data Fidelity
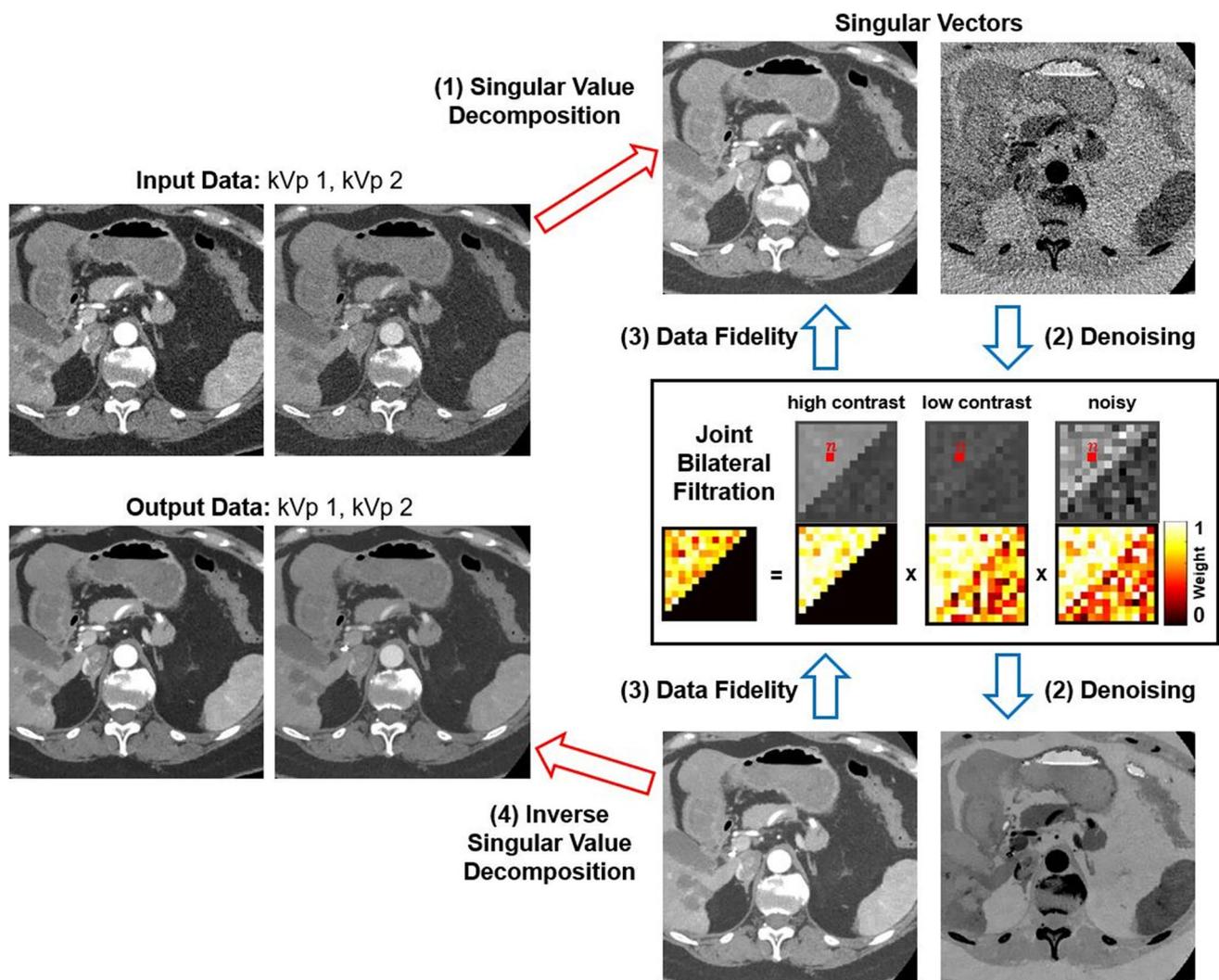
(2) Denoising

(4) Inverse Singular Value Decomposition

Fig. 2 Rank-sparse kernel regression (RSKR) is a denoising algorithm designed for multi-energy CT data. (1) Example weighted singular value decomposition of dual-energy data which yields a high-fidelity average image and a lower fidelity difference image. (2–3) Denoising is performed by applying joint bilateral filtration, an edge preserving smoothing operation which matches edges between the average and difference images. Step 2 is alternated with fidelity updates (3), to prevent loss of image details, and further iterated as needed. (4) Denoised results following inversion of the singular value decomposition

not accounting for changes in in-plane-pixel size. The MTF curve provides a measure of how well the system transfers the contrast ratio of the original object (y-axis) across spatial frequencies (x-axis).

2D noise power spectrum (NPS) measurements were taken to comprehensively assess noise properties. Measurements were made in 32^2 voxel homogenous regions of interest in subcutaneous fat to assess for objective noise levels. The NPS curve shows the variance in image intensity over spatial frequencies in an image. MTF and NPS measurements were averaged over a subset of 18 patients with matching reconstruction parameters (Flash scanner, 1.5 mm slice thickness with 0.98 mm in-plane-voxel size) using the same measurement regions between algorithms.

Qualitative image quality analysis

Reference standard

To enable randomized and blinded reading, one representative slice from each examination was chosen based on the main diagnosis in the impression of the clinical radiology report. Where exact slice positions were dictated, these were used for slice selection. In cases without a dictated slice position, one radiology fellow (F.R.S., 9 years of experience) selected the slice that showed the main diagnosis best, based on the standard clinical image reconstruction. In cases without pathology, a slice was chosen at the porta hepatis. Each slice was reconstructed with the regular clinical

reconstruction parameters, and denoised with the comparison algorithm and the test algorithm. The slice window level was adjusted to produce the most similar visual impression between slices (most used: 300/50 and 650/225) and readers were not able to adjust these during their assessment.

Readers

Slices were presented to four readers not involved in image selection (5, 7, 8, and 9 years of experience) in randomized order in three separate sets that each included all denoising conditions. Each patient was represented once per set and readers completed one set at a time, coming back to the same set whenever needed. They were not allowed to change scores on a previous image set, once it had been completed.

Readers were provided with a reference document for scoring. They were asked to provide scores from 0 to 100 (higher score = better evaluation) for image noise, general image quality, and their comfort in making a diagnosis based on the image provided.

In addition, readers were asked to make a diagnosis based on a table with 80 possible diagnoses (Table 3, adapted from Jaffe et al [11]), which included all diagnoses from the imaging reports. More than one diagnosis was counted as correct, if they were mentioned in the report and visible on the image. Diagnoses made by readers were counted as correct, if at least one of the diagnoses they entered was correct and they were forced to make a diagnosis before they could complete the reads. Overall, there were 948 reads (79 patient cases * 3 denoising conditions * 4 readers).
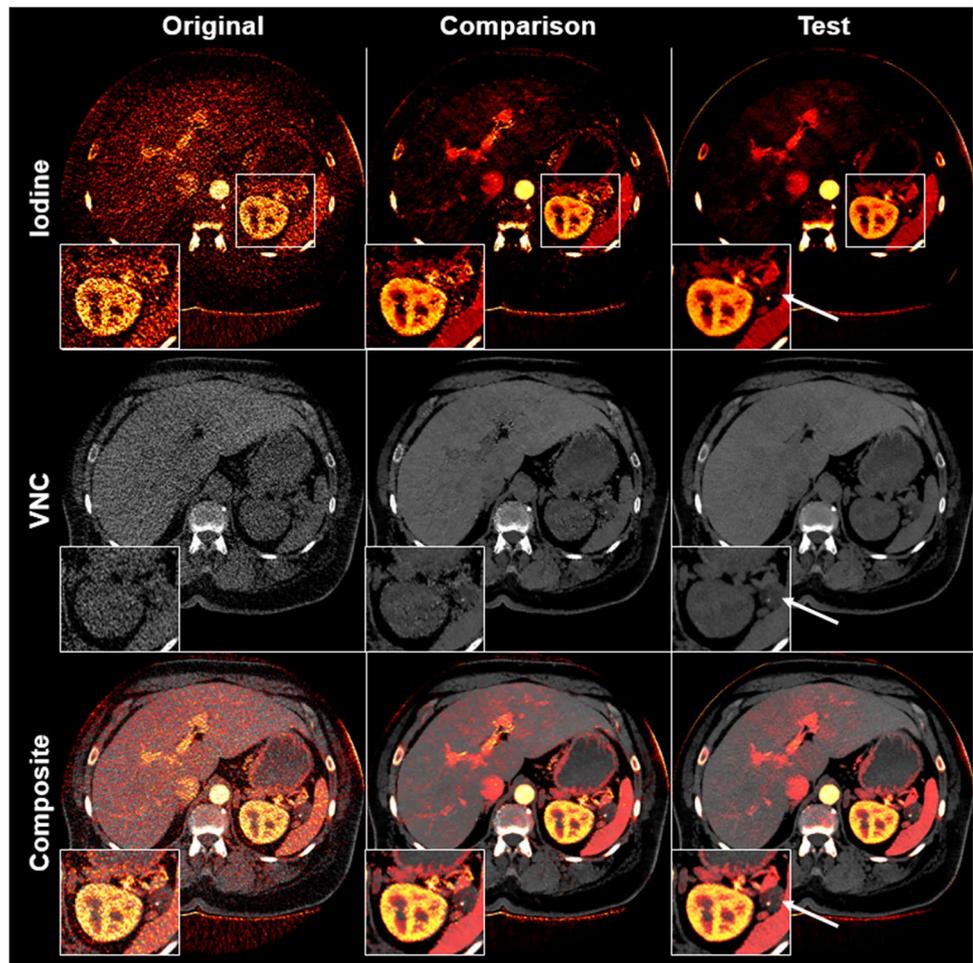
Statistical analysis

Descriptive statistics were used to characterize the patient population. Image scores were compared using paired two-tailed t-tests. Reader accuracy was compared using one-way ANOVA with Bonferroni correction. Reader scores were compared per denoising algorithm and the assumptions for linear models were confirmed. To avoid biasing the results with clustering by readers or patients, scores were compared between algorithms using mixed-effects linear regression with denoising algorithm treated as a fixed effect and patient and reader as random effects. Post hoc testing was performed using the Tukey test. Inter-reader agreement was determined using intraclass correlation coefficients (ICC) and Spearman's rank correlation test. The accuracy of diagnosis was determined based on the reference standard. The level of statistical significance was assumed at alpha = 0.05. Statistical analysis was performed using RStudio Version 1.3.1056.

Table 3 The diagnoses that were options for the readers to choose from. 0 was considered as "no pathology"

Organ system	Diagnosis A	Diagnosis B	Diagnosis C	Diagnosis D	Diagnosis E
Lung bases	Atelectasis (1)	Consolidation (2)	Effusion (3)	Nodule (4)	Fibrosis (5)
Cardiac	Cardiomegaly (6)	Effusion (7)	Coronary Calcium (8)	Pacemaker (9)	Enlarged (10)
Liver	Mass (11)	Fat (12)	Enlarged (13)	Cirrhosis (14)	Ascites (15)
Gallbladder	Absent (16)	Stone/Sludge (17)	Wall thickening (18)	Fluid (19)	Mass (20)
Biliary tree	Duct dilatation (21)	Stone (22)	Mass (23)	Pneumobilia (24)	/
Pancreas	Mass (25)	Stranding (26)	Duct dilatation (27)	Calcification (28)	Fluid (29)
Spleen	Mass (30)	Enlarged (31)	Calcification (32)	Infarct (33)	Splenule (34)
Adrenals	Mass (35)	Calcification (36)	Hyperplasia (37)	/	/
Kidneys	Mass (38)	Stone (39)	Hydronephrosis (40)	Pyelonephritis (41)	Infarct (42)
Ureters	Mass (43)	Stone (44)	Dilatation (45)	/	/
Stomach	Mass (46)	Wall thickening (47)	Ulcer (48)	Distension (49)	/
Small Bowel	Mass (50)	Wall thickening (51)	Ileus (52)	Dilatation (53)	Obstruction (54)
Colon	Mass (55)	Wall thickening (56)	Diverticulosis (57)	Diverticulitis (58)	Appendicitis (59)
Lymph nodes	Increase size (60)	Increase number (61)	Calcification (62)	/	/
Arteries	Aneurysm (63)	Thrombosis (64)	Calcification (65)	Variants (66)	Stent (67)
Veins	Varices (68)	Thrombosis (69)	Calcification (70)	Variants (71)	Stent (72)
Bones	Mass (73)	Infection (74)	Fracture (75)	/	/
Others	Surgical changes (76)	Ovarian mass (77)	Fat stranding (78)	Atrophy (79)	Hernia (80)

Fig. 3 Dual-energy post-processing into iodine and virtual non-contrast (VNC) images for a 55-year-old female patient with a pancreatic tail lesion. Insets highlight improved iodine map image quality following denoising with the test algorithm and the location of the lesion (white arrows)

Results

Quantitative image quality analysis

Average CNR values were 5.0 ± 1.9 (original), 31.1 ± 10.3 (comparison; $p < 0.001$), and 8.9 ± 2.9 (test; $p < 0.001$). These quantitative measures are correlated with reader perceptions of noise detailed in the next section. The denoising algorithms did not aversely affect spatial resolution measurements at the skin-air boundary (MTF at 50%: ~ 3.0 cm^{-1}; MTF at 10%: ~ 5.2 cm^{-1} for all algorithms). Both denoising algorithms significantly reduced noise power relative to the original reconstructions at high spatial frequencies and performed well at the creation of iodine maps and virtual non-contrast images (an example is shown in Fig. 3). The ME-NLM algorithm removed more noise at mid spatial frequencies than RSKR (Supplemental Fig. 1).

Qualitative image quality analysis

The perceived image noise received a mean score of 54 ± 21 (original), 84 ± 10 (comparison), and 66 ± 17 (test).

The difference in perceived image noise was significant between algorithms with the comparison algorithm receiving an average of 29.6 higher scores than the original images ($p = 0.045$) and the test algorithm receiving an average of 11.8 higher scores than the original images ($p = 0.014$; Table 4). There was no statistically significant difference between the scores for the comparison algorithm and the test algorithm ($p = 0.102$). Readers were overall in good agreement over perceived image noise (ICC: 0.83). This finding

Table 4 Mean scores for perceived image noise, image quality, and reader confidence in making a diagnosis with 95% confidence intervals; bold * statistically significant improvement over standard reconstruction

	Original	Comparison	Test
Image noise	54.0 [21.0–87.0]	**83.6*** [68.4–98.7]	**65.8*** [34.9–96.7]
Image quality	59.4 [33.7–85.1]	61.1 [23.9–98.3]	64.5 [34.8–94.3]
Diagnostic comfort	63.4 [38.5–88.4]	67.5 [34.5–100.5]	67.8 [37.0–98.6]

correlates with the quantitative measures of reduced image noise for both of the denoising algorithms, the perception of the readers agreeing with the NPS measurements.

The perceived image quality received a mean score of 59 ± 16 (original), 61 ± 24 (comparison), and 65 ± 19 (test). The difference in perceived overall image quality was not statistically significant for the comparison algorithm (average of 1.7 higher, $p = 0.935$) or the test algorithm (average of 5.1 higher, $p = 0.293$). There was no statistically significant difference between the scores for the comparison algorithm and the test algorithm ($p = 0.480$). Readers were in moderate agreement over image quality (ICC: 0.72).

The diagnostic comfort received a mean score of 64 ± 16 (original), 68 ± 21 (comparison), and 68 ± 20 (test). The difference in diagnostic comfort was not statistically significant for the comparison algorithm (average of 4.1 higher, $p = 0.489$) or the test algorithm (average of 4.3 higher, $p = 0.351$). There was no statistically significant difference in diagnostic comfort between the scores for the comparison algorithm and the test algorithm ($p = 0.986$). Readers were in moderate agreement over diagnostic comfort (ICC: 0.66).

Diagnostic accuracy, where readers gave at least one of the diagnoses mentioned in the report, was relatively low across algorithms (accuracy: 66, 63, and 67% for original, comparison, and test; Fig. 4). These differences were not statistically significant when comparing accuracy between original and comparison algorithm ($p = 0.527$), original and test algorithm ($p = 0.741$), or the two different denoising algorithms ($p = 0.202$).

There were differences between individual readers in preference for algorithm and the diagnostic accuracy depending on algorithm and reader (Fig. 5, Supplemental Figs. 2 and 3). Two readers gave the highest diagnostic comfort score to the RSKR, but one showed better diagnostic accuracy on the original images, while one other reader gave the highest diagnostic comfort score to the original images but had the highest diagnostic accuracy with our test algorithm (Table 5, Fig. 6).

Discussion

Our study showed that diagnostic accuracy was overall low for obese patients and only slightly higher when using our denoising algorithm. The average scores for image noise, image quality, and diagnostic comfort were higher than for the original iterative reconstructions without denoising,
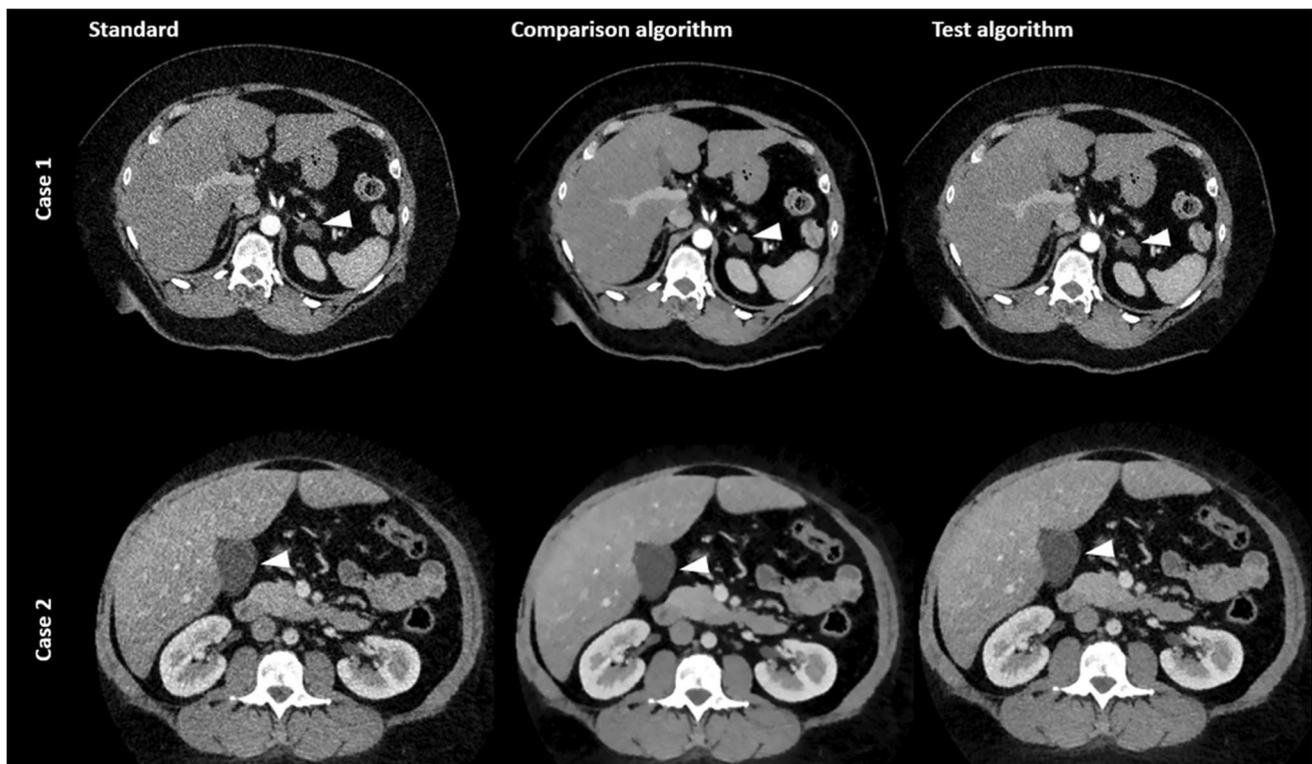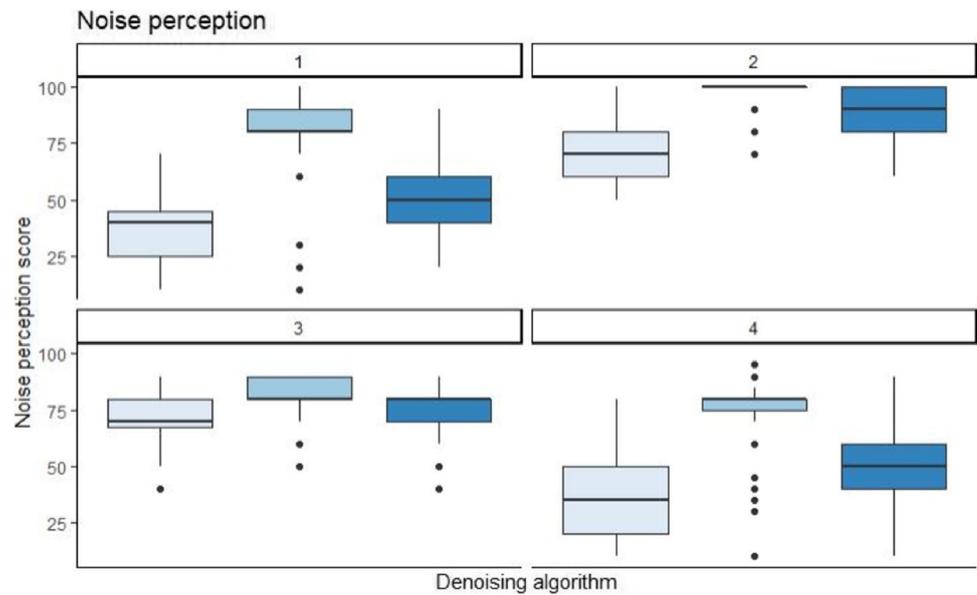

Fig. 4 Two challenging cases. Case 1 (top row), in which only two of the readers made the diagnosis of left adrenal lesion in a 43 year-old female patient with a BMI of 41.3 kg/m^2 on the original reconstruction with standard parameters. Case 2 (bottom row), in which only two of the readers made the diagnosis of thickened gall bladder wall in a 37-year-old female patient with a BMI of 36.2 kg/m^2 on the original reconstruction with standard parameters

Fig. 5 The differences in image noise perception by reader and denoising algorithm. The scores for original images are shown in pale blue (left), for the comparison algorithm in light blue (middle), and for the test algorithm in dark blue (right)

and the highest image quality scores were given to the RSKR algorithm. This is similar to results reported for commercially available denoising algorithms applied to single-energy, low-dose CT, where image quality was rated as poor with FBP but moderate when using the denoising algorithm [12].

The comparison algorithm used received the best noise scores, but lower average subjective image quality scores than the test algorithm (61 vs. 65). This may have been due to the fact that the algorithm smoothes out the image texture to a point where radiologists are uncomfortable with the "plastic" appearance of abdominal structures, which has been described for high iterative-reconstruction strengths [13, 14] and may in part be due to changes in NPS [10].

There was some disagreement between which of the denoised results the radiologists preferred and how accurate they were in making their diagnoses. Two readers preferred the RSKR but one was less accurate in their diagnoses with RSKR than with the original reconstruction. One other reader preferred the ME-NLM results but was the most accurate with RSKR. While this is an unexpected effect and we did not record time spent interpreting each image, it might be due to readers spending more time scrutinizing the images they preferred less and thus reaching the correct diagnosis more often. The effect of time spent looking at images and diagnostic accuracy has been shown for mammography data [15], but should be explored further in future studies.

In addition, though there was a general preference for denoised images of any kind, both readers with the highest comfort scores assigned to RSKR were trained at our institution, while the readers who preferred ME-NLM received their initial training at a different institution. There might

be an effect of diagnostic comfort being dictated by an image that is closer to what the training institution was using (e.g., iterative reconstruction vs. filtered back projection as default), since protocols vary widely between institutions nationally and internationally [16, 17].

Recently, radiology reader fatigue has come to be a topic of some research interest and it might be reasonable to assume that looking at images of higher quality could

Table 5 The individual scores given by readers for image noise, quality, and diagnostic comfort. Highlighted are the best scores per category for each reader

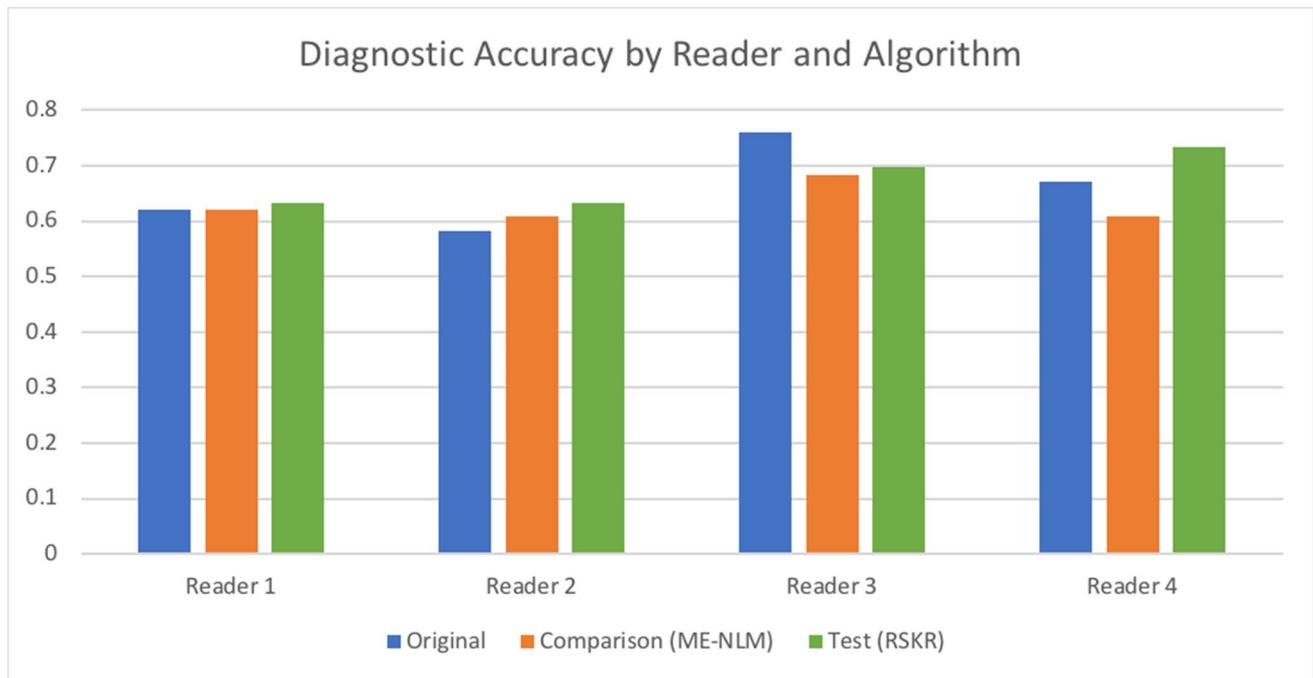|	Reader 1	Reader 2	Reader 3	Reader 4
Image noise Original	35.6±13.2	73.0±12.2	71.1±10.1	36.2±18.2
Image noise Comparison	**77.5±17.7**	**97.3±5.9**	**83.4±8.1**	**76.1±13.9**
Image noise Test	49.9±16.3	87.5±12.4	77.4±9.3	48.5±19.5
Image quality Original	40.5±14.4	76.9±13.1	68.5±9.9	51.7±20.4
Image quality Comparison	39.2±15.8	**92.8±10.0**	64.0±7.7	48.5±19.7
Image quality Test	**47.5±18.6**	89.0±11.2	**69.8±9.3**	**52.0±21.2**
Diagnostic comfort Original	49.3±25.7	79.6±20.9	75.1±16.2	**49.7±29.8**
Diagnostic comfort Comparison	52.0±25.5	**92.7±12.0**	77.1±14.7	48.3±31.5
Diagnostic comfort Test	**56.2±26.7**	89.1±14.5	**79.0±15.2**	47.0±33.6

Fig. 6 The inter-reader differences in diagnostic accuracy based on whether and which form of denoising was applied to the images. Readers 1 and 3 preferred the images generated by the test algorithm but only reader 1 also had highest accuracy using these images, while readers 2 and 4 preferred the overall image quality from the comparison algorithm but had highest accuracy with the test algorithm

lead to less reader fatigue over the course of a work day even if it does not improve accuracy on a per-scan basis, which should be evaluated in future studies [18].

Though overall diagnostic accuracy was not statistically significantly improved, using denoising algorithms might be a way of reducing radiation dose in obese patients, which should be explored in prospectively designed future studies. Since this denoising technique was added to the standard iterative reconstruction already used in clinical practice, it would also be important to integrate these techniques into the background workflows performed at the scanner or in PACS so using them will not lead to increased workloads for technologists or radiologists. This would also increase the availability of the denoising techniques to obese patients in all of the USA and potentially worldwide, instead of it being used only at highly specialized academic centers.

Limitations

In addition to the retrospective nature, several limitations of our study deserve attention. First our study was limited by the fact that it was a single-center study with a relatively small sample size and with variance in the scanner types and acquisition protocols due to the challenges of acquiring CT data from obese patients. Second, additional improvement in diagnostic quality might be achieved by providing readers with the full dataset and clinical history, which should be

explored in follow-up investigations, since the task of making a diagnosis based on a single CT slice is particularly difficult and may have contributed to the overall low diagnostic accuracy in our study. Third, the retrospective nature of data acquisition does not provide the opportunity to draw direct conclusions about clinical management; this should be further investigated in a prospective study designed to test clinical performance of denoising algorithms in obese patients. The reader study was limited to a comparison between the standard iterative reconstructions (ADMIRE/SAFIRE 2) and a non-local-means based algorithm (ME-NLM). In the future, it might be feasible to broaden the comparisons to deep learning algorithms and potentially evaluate denoising for virtual monoenergetic imaging.

For future directions, since our algorithm works in the image domain and is not specific to an acquisition method, we believe it would be applicable for the emerging technology of photon-counting CT data, which may also profit from denoising algorithms. In a recent study, the RSKR denoising algorithm was applied during iterative reconstruction to compare preclinical micro-CT data using energy-integrating detectors and a photon-counting detector with four energy thresholds to image iodine and gadolinium based contrast agents in sarcoma tumors induced on the flanks of mice [19]. The combination of inherently higher contrast resolution in photon-counting CT and denoising algorithms could further improve imaging of obese patients.

In conclusion, the new algorithm produces quantitatively higher image quality data than current standard and existing denoising algorithms in obese patients imaged with DECT. Readers show a preference for it and their average diagnostic accuracy has a tendency for improvement.

Declarations

Guarantor The scientific guarantor of this publication is Daniele Marin.

Conflict of interest Fides R. Schwartz and Daniele Marin – speaker honoraria from Siemens Healthineers.

Fides R. Schwartz is a member of the European Radiology Editorial Board. They have not taken part in the review or selection process of this article.

The other 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 This was a retrospective study and the IRB waived the need for informed consent from individual patients. No animal subjects were studied.

Ethical approval Institutional Review Board approval was obtained prior to the start of the study.

Methodology
- retrospective
- cross-sectional
- performed at one institution

References

1. Megibow AJ, Sahani D (2012) Best practice: implementation and use of abdominal dual-energy CT in routine patient care. AJR Am J Roentgenol 199:S71–S77
2. Modica MJ, Kanal KM, Gunn ML (2011) The obese emergency patient: imaging challenges and solutions. Radiographics 31:811–823
3. Uppot RN (2018) Technical challenges of imaging & image-guided interventions in obese patients. Br J Radiol 91:20170931
4. Rubin GD (2000) Data explosion: the challenge of multidetector-row CT. Eur J Radiol 36:74–80
5. Langer SG (2011) Challenges for data storage in medical imaging research. J Digit Imaging 24:203–207
6. Mileto A, Guimaraes LS, McCollough CH, Fletcher JG, Yu L (2019) State of the art in abdominal CT: the limits of iterative reconstruction algorithms. Radiology 293:491–503
7. Clark DP, Badea CT (2017) Hybrid spectral CT reconstruction. PLoS One 12:e0180324
8. Li Z, Yu L, Trzasko JD et al (2014) Adaptive nonlocal means filtering based on local noise level for CT denoising. Med Phys 41:011908
9. Li Z, Leng S, Yu L, Manduca A, McCollough CH (2017) An effective noise reduction method for multi-energy CT images that exploit spatio-spectral features. Med Phys 44:1610–1623
10. Greffier J, Frandon J, Larbi A, Beregi J, Pereira F (2020) CT iterative reconstruction algorithms: a task-based image quality assessment. Eur Radiol 30:487–500
11. Jaffe TA, Martin LC, Miller CM et al (2007) Abdominal pain: coronal reformations from isotropic voxels with 16-section CT–reader lesion detection and interpretation time. Radiology 242:175–181
12. Steuwe A, Weber M, Bethge OT et al (2021) Influence of a novel deep-learning based reconstruction software on the objective and subjective image quality in low-dose abdominal computed tomography. Br J Radiol 94:20200677
13. Mohammadinejad P, Mileto A, Yu L et al (2021) CT noise-reduction methods for lower-dose scanning: strengths and weaknesses of iterative reconstruction algorithms and new techniques. Radiographics 41:1493–1508
14. Geyer LL, Schoepf UJ, Meinel FG et al (2015) State of the art: iterative CT reconstruction techniques. Radiology 276:339–357
15. Carney PA, Bogart TA, Geller BM et al (2012) Association between time spent interpreting, level of confidence, and accuracy of screening mammography. AJR Am J Roentgenol 198:970–978
16. Willatt JMG, Mason AC (2006) Comparison of radiology residency programs in ten countries. Eur Radiol 16:437–444
17. Homayounieh F, Holmberg O, Umairi RA 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
18. Taylor-Phillips S, Stinton C (2019) Fatigue in radiology: a fertile area for future research. Br J Radiol 92:20190043
19. Badea CT, Clark DP, Holbrook M, Srivastava M, Mowery Y, Ghaghada KB (2019) Functional imaging of tumor vasculature using iodine and gadolinium-based nanoparticle contrast agents: a comparison of spectral micro-CT using energy integrating and photon counting detectors. Phys Med Biol 64:065007