

Correlation of Automated *in Vivo* Image Quality With Radiologist's Performance in Abdomen Computed Tomography Across Conventional and Deep Learning Reconstructions

Mojtaba Zarei, PhD,* Francesco Ria, PhD,† Corey T. Jensen, MD,‡
Xinning Liu, PhD,§ Craig K. Abbey, PhD,|| and Ehsan Samei, PhD¶

Objective: Image quality evaluation in radiology is most relevant when reflects radiologists' performance. This study assessed how image quality measurement in terms of *in vivo*-characterized detectability index (d') for low-contrast liver lesion assessment in CT is correlated with radiologists' performance across 2 different CT reconstructions.

Methods: Fifty-one contrast-enhanced abdominal studies for investigating colorectal liver metastases were prospectively performed using 2 radiation dose exposures and reconstructed with Filtered back projection (FBP) and deep learning image reconstruction (DL) algorithms for a total of 161 noncalcified hypoattenuating lesions for 3 lesion size (D) subsets (<6 mm, 6 to 10 mm, and >10 mm). Images were assessed by expert radiologists for hepatic lesion detection task and likelihood of malignancy across the 2 imaging conditions. All cases were also evaluated automatically in terms of *in vivo* d' as a metric of task-based performance, both using a conventional technique and a new formalism of an added frequency term in the internal noise component of d' to accommodate the nonlinearity of the DL reconstruction (d' adj).

Results: The study found conventionally defined d' well-reflective of radiologists' evaluation of FBP images but not well-aligned with that of DL images. The new formalism provided more consistent reflection of performance across reconstruction techniques. In particular, in the lesion group $D \leq 6$ mm, the difference between radiologists' accuracy in images acquired with DL and images acquired with FBP was -26%, and the related d' adj difference was -9%, whereas the d' was 34%. Analogously, for the lesion group $6 \text{ mm} < D \leq 10$ mm, the differences were -15%, -13%, and 29%, respectively. Lastly, for the lesion group $D > 10$ mm, radiologists showed the same accuracy in both FBP and DL images, difference in d' adj was -11%, and difference in d' was 31%.

Conclusion: The new d' formalism can robustly reflect CT systems clinical performance irrespective of reconstruction algorithm. The methodology can be more readily applied to assess the real-world performance of CT systems.

Key Words: detectability index, radiologists' performance, clinical CT performance, deep learning reconstruction

(*J Comput Assist Tomogr* 2026;00:000-000)

Adequate image quality at the lowest radiation risk to patient is the overall goal of every CT acquisition and the goal of technical improvements in CT. Such improvements can be introduced in both the scanner's hardware and software. Examples of the first category are the state-of-the-art new acquisition technologies, such as photon counting CT.^{1,2} Of the second are the new reconstruction techniques. In particular, the implementation of iterative reconstruction (IR) and deep learning image reconstruction (DL) methods promise improved image quality and reduced radiation dose. However, to assess such potential, there is a need to assess image quality *in vivo* in such a way that is reflective of the actual clinical performance of the system.^{3,4}

Over the years, several approaches have been introduced to objectively measure image quality from patient images in terms of metrics such as noise magnitude, spatial resolution, and image Hounsfield Units-contrast characteristic and distribution.⁵⁻⁷ Spatial resolution and noise quantities can be combined to derive an overall metric of image quality, the so-called detectability index (d'), which represents the combined effect of those metrics to represent

Received for publication September 2, 2025; accepted December 3, 2025.

From the *Department of Radiology, Duke University Health System, Center for Virtual Imaging Trials and Carl E. Ravin Advanced Imaging Labs, Duke University Health System, Durham, NC; †Department of Radiology, Duke University Health System, Center for Virtual Imaging Trials, Clinical Imaging Physics Group, and Carl E. Ravin Advanced Imaging Labs, Durham, NC; Department of ‡Abdominal Radiology; §Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX; ||Department of Psychological and Brain Sciences University of California Santa Barbara, Santa Barbara, CA; and ¶Departments of Radiology, Physics, Biomedical Engineering, and Electrical and Computer Engineering, Duke University Health System, Center for Virtual Imaging Trials, Clinical Imaging Physics Group, Carl E. Ravin Advanced Imaging Labs, Medical Physics Graduate Program, Durham, NC.

F.R. and M.Z. equally contributed to this work.

This work was funded in part by NIH/NIBIB P41EB028744 and in part by NIH/NIBIB R44EB031658.

E.S. discloses relationship with the following entities unrelated to the present publication: GE, Siemens, Bracco, Imaloxig, 12Sigma, Sun-Nuclear, Metis Health Analytics, Cambridge University Press, and Wiley and Sons. F.R. discloses relationship with Metis Health Analytics, unrelated to the present publication. C.J. received in-kind grants from General Electric for the evaluation of deep-learning reconstructions in CT unrelated to the present publication. C.K.A. discloses relationship with the following entities unrelated to the present publication: GE, Canon Medical Research, and Izotropic LLC.

Correspondence to: Francesco Ria, PhD, Department of Radiology, Duke University Health System, Center for Virtual Imaging Trials, Clinical Imaging Physics Group, and Carl E. Ravin Advanced Imaging Labs, 2424 Erwin Road, Suite 302, Durham, NC 27710 (e-mail: francesco.ria@duke.edu).

Copyright © 2026 Wolters Kluwer Health, Inc. All rights reserved.

DOI: 10.1097/RCT.0000000000001845

a specific task based on signal detection theory.^{8,9} The detectability index represents a generic reflection of the information content of images, applicable to both detection and classification tasks.¹⁰ However, d' is relevant only to the extent that it reflects the radiologists' interpretation of images. This is particularly relevant to reconstruction techniques, including deep learning methods, which use nonlinear processes.¹¹⁻¹³ Can d' reflect the radiologists performance as it varies across reconstruction methods?

The purpose of this study was to assess *in vivo* d' for 2 different CT reconstruction techniques at 2 dose levels, and to compare it with radiologist's performance in the detection and characterization of low-contrast liver lesions in abdominal CT. In particular, contrast-enhanced abdominal studies for investigating colorectal liver metastases were prospectively performed using 2 radiation dose exposures and reconstructed with Filtered back projection (FBP) and DL algorithms. The interpretation accuracy was then compared with 2 different d' computational methods.^{14,15}

MATERIALS AND METHODS

Clinical CT Image Dataset and Qualitative Analysis

The clinical images were obtained from a prior prospective patient study¹¹ with Institutional Review Board approval. The study evaluated Deep Learning (DL)-based reconstruction and FBP reconstruction contrast-enhanced CT examinations of the abdomen (performed in the same breath-hold) for the detection of hepatic metastases in patients with colorectal adenocarcinoma. The study included clinical CT scans of 51 patients reconstructed under 2 different imaging conditions for a total of 102 image datasets. In particular, the imaging conditions included: (1) a standard dose (SD) and a 65% reduced dose acquisition (RD); and (2) 2 different CT reconstructions, including a standard reconstruction with Filtered Back Projection (FBP), and DL-based medium-strength Reconstruction (DL-medium). The imaging protocol and radiation dose parameters used for the clinical CT image dataset are listed in Table 1.

The evaluation also included assessments of radiologist's accuracy in assessing lesion's malignancy for 3 lesion size subsets (<6 mm, 6-10 mm, and >10 mm) and 2 imaging conditions (ie, standard dose FBP, and reduced dose DL-based kernel) on a total of 161 noncalcified hypoattenuating lesions (127 metastatic and 34 benign, 2 to

15 mm in diameter). Details of the analysis and the related results are reported in a previous study.¹¹ In particular, 3 radiologists used a Likert-type scores for each lesion with respect to characterization (1 = definitely benign; 5 = definitely malignant) and related confidence in the diagnosis (1 = lowest confidence; 5 = highest confidence). The ground truth was established by full agreement of 2 nonblinded consensus reviewers.

Quantitative Analysis

Conventional Observer Model

All cases were processed using a validated algorithm for automated *in vivo* image quality metrics to extract a d' based on the Fisher-Hotelling (FH) observer model with an eye filter (E) and internal noise (N) for 3 categories of lesion sizes: smaller than 6 mm, between 6 and 10 mm, and greater than 10 mm. This conventional d' was formulated as¹⁵

$$d'^2 = \int \frac{W^2 MTF^2 E^2}{E^2 NPS(1 + \beta) + N} dudv, \quad (1)$$

where W represents the mathematical model of the considered task incorporating lesion size and contrast, observer distance to monitor, the field of view, and screen size; β is a non-negative coefficient, MTF is the modulation transfer function computed over the air-skin region in CT scans, and NPS is the noise power spectrum measured over a uniform region inside the body.^{6,16,17} Following Richard and Siewerdsen,¹⁸ N was calculated as

$$N = \alpha D^2 \sigma, \quad (2)$$

where D is the reader distance to the monitor, σ is the SD of the noise in the image, and α is a small positive coefficient ($0 < \alpha \leq 1$) determined empirically.

Adjusted Observer Model

Some prior studies have shown that Equation (2) may not fully capture the human perception of the noise,¹⁹⁻²² particularly given the frequency-dependent filtering of DL reconstruction methods.²³ As such we generalized the internal noise formulation with an added frequency component as

$$N = \alpha D^2 \sigma_f^\gamma, \quad (3)$$

TABLE 1. Scanner Parameters for the Acquisition of the Clinical Dataset

Parameter	Value
Scanner vendor and model	GE health care—revolution CT ES
Detector configuration (mm)	128 × 0.625
Beam collimation (mm)	80
Acquisition section thickness (mm)	5, full mode
Reconstruction thickness and increment (mm)	2.5/2.5, plus mode
Pitch	0.508:1
Rotation time (s)	0.5-0.7
Table speed (mm/rotation)	40.64
Radiation dose requirements:	SD: settings to approximate standard exam. RD: settings to approximate 65% dose reduction.
Tube potential (kV)	120
Image reconstruction algorithms	FBP: filtered back projection Medium-DL: medium strength deep learning image reconstruction
Reconstruction kernel	Standard plus

where γ is a nonpositive number ($\gamma \leq 0$), and f_r is the radial frequency. By this formulation, instead of assuming a constant internal noise across all frequency, the model allows the internal noise to be lower at higher frequencies, consistent with high-contrast visual-contrast sensitivity. In the implementation, to avoid singularity, a small positive constant was added to f_r . The resulting equation for the adjusted d' was thus formulated as

$$d'_{adj} = \int \frac{W^2 MTF^2 E^2}{E^2 NPS(1 + \beta) + \alpha D^2 \sigma_f^\gamma} dudv. \quad (4)$$

To assess how the d' values correlate with radiologists' results, a heuristic optimization problem using genetic algorithm (GA) approach was solved to select the hyperparameters α , β , and γ . The initial step involved the definition of decision variable bounds, specifying the permissible ranges for each hyperparameter. The boundary of the search for each of the hyperparameters was set to

$$10^{-8} \leq \alpha \leq 1,$$

$$0 \leq \beta \leq 1,$$

and

$$-3 \leq \gamma \leq 0.$$

The hyperparameters were used to calculate d' values for both FBP and DL reconstructed cases. Assuming DL condition may only influence the frequency component of the internal noise, α and β were considered common parameters across both reconstructions (reflecting conventional d' methodology), but γ was allowed to be reconstruction dependent. For each set, the resultant d' for FBP and DL-based kernel were evaluated at each lesion size category.

The optimal set of hyperparameters was constrained to ensure that increased or decreased d' values correspond with similar trend in the accuracy of the radiologists' readings for each lesion size category. Set up as an optimization problem, the goal was to minimize the deviation between radiologists results and d' results across lesions sizes. To prevent zero nominators, we used the logarithm of the radiologist's accuracy rate. The objective value was to minimize the SD of the absolute deviation between accuracy rate difference and d' difference across all lesion sizes.

To conduct the optimization, the distributed evolutionary algorithm Python library (DEAP) was used.²⁴ An initial population of 500 candidate solutions was initiated, with random parameter values falling within predefined bounds. The NSGA-II (Non-dominated Sorting Genetic Algorithm II) selection strategy²⁵ was utilized to determine which solutions should serve as parents for the subsequent generation. A blending strategy was applied during the

crossover operation, where parent solutions were paired, and their genetic material was combined. Therefore, during the crossover operation, the genetic material of the offspring was a balanced blend of the genetic material from both parent individuals. To promote exploration of the solution space, a Gaussian mutation with a zero mean and 0.2 SD was implemented. Furthermore, the initial mutation rate was set at 0.2 and dynamically increased from 0.25 to 1 after the fifth generation. For each set of candidates, feasibility and objective values were assessed, and candidates failing to meet the constraints were removed from the population. The optimization procedure was iterated for 100 generations, and the candidates exhibiting the lowest objective value within the feasible domain were recorded as the optimal hyperparameters. Radiologists' accuracy, d' , and d'_{adj} were compared qualitatively and quantitatively terms of percentage differences between FBP and DL images.

RESULTS

The median accuracy for radiologists for the 3 nodule size categories, the d' , and d'_{adj} values for both FBP and DL images are reported in Table 2, together with the related SDs. Figure 1 depicts the plot of d' versus the median accuracy by radiologists for 3 nodule size categories before adjustment. While there is an increase in both d' and accuracy rate as the nodule size increases, the changes in accuracy and d' rates are not consistent in the FBP and DL images. For example, for small- and medium-sized lesions, the radiologist accuracy rate for FBP is higher than the radiologist accuracy rate in DL-based images. However, d' in FBP was smaller than the d' in DL images, which was inconsistent with radiologists' results.

The desired condition for the hyperparameters was to find common hyperparameters for both FBP and DL-based images. However, excessive domain search did not produce a feasible solution, except when γ was allowed to be different for FBP and DL images, thus d'_{adj} could be calculated. The optimal parameters were $\alpha = 0.25$ and $\beta = 0.87$ for both FBP and DL images and $\gamma = -2.8$ for FBP, and $\gamma = -2.4$ for DL images. The radiologist accuracy rate versus the d'_{adj} is depicted in Figure 2. The plot shows higher values from DL images to FBP images for the same lesion size category, aligned with radiologists' accuracy rate.

Figure 3 shows the distribution of the d' for the default parameters, radiologist accuracy rate variability, and d'_{adj} distribution across 3 lesion size categories. It shows that the new distribution of the detectability index better represents the accuracy rate distribution. Across all lesion size categories, unlike conventional d' , d'_{adj} had consistently lower values for the DL images compared with the FBP images, which aligns with the radiologists' evaluation.

As showed in Figure 4, in the lesion group D < = 6 mm, the difference between radiologists' accuracy in

TABLE 2. For Each Lesion Group Size: Median Radiologist Accuracy, d' , and d'_{adj} Values for FBP and DL Images, and Related SDs

Lesion Group	FBP			DL		
	Radiologist Accuracy (%)	d'	d'_{adj}	Radiologist Accuracy (%)	d'	d'_{adj}
D < = 6 mm	47.69 ± 50.34	1.07 ± 0.51	0.11 ± 0.05	35.38 ± 48.19	1.43 ± 0.67	0.10 ± 0.04
6 mm < D < = 10 mm	91.67 ± 27.83	2.96 ± 0.92	0.30 ± 0.09	77.78 ± 41.87	3.81 ± 1.21	0.26 ± 0.08
D > 10 mm	95.83 ± 20.41	6.08 ± 1.36	0.61 ± 0.10	95.83 ± 20.41	7.98 ± 1.78	0.54 ± 0.09

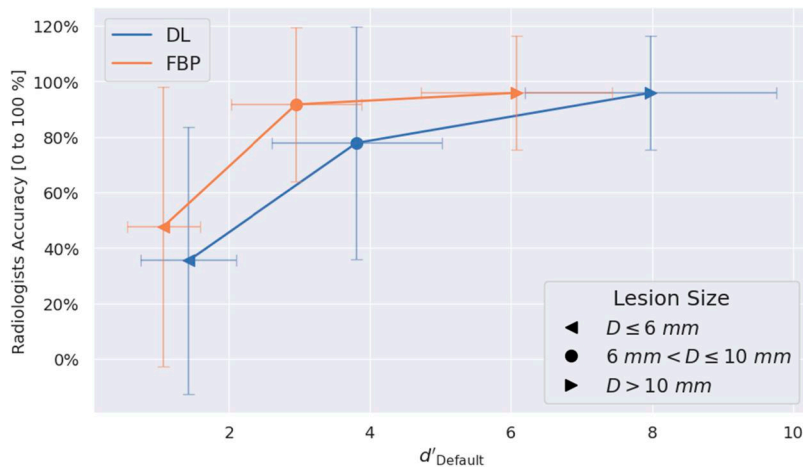


FIGURE 1. Original d' values versus radiologist’s accuracy. Error bars are the calculated SDs. For the small- and medium-sized lesions when we switch FBP to DL-based, despite the decrease in the radiologist’s accuracy, d' value increases.

images acquired with DL and images acquired with FBP was -26% , the related d'_{adj} difference was -9% , whereas the d' was 34% . For the lesion group 6 mm < $D \leq 10$ mm, the difference between radiologist accuracy in images acquired with DL and images acquired with FBP was -15% , the related d'_{adj} difference was -13% , whereas the d' was 29% . Lastly, for the lesion group $D > 10$ mm, radiologists showed the same accuracy in both FBP and DL images, difference in d'_{adj} was -11% , and difference in d' was 31% .

DISCUSSION

Assessment of image quality from patient images provides a significant insight into the evaluation and optimization of clinical imaging performance. However, such methodology is valid only if it produces results consistent with clinical performance. In this study, we evaluated this goal in the context of standard and deep

learning-based reconstructions. Two different calculation strategies for the *in vivo* detectability index (d') were compared with radiologists’ performance in the assessment of low-contrast lesions in abdominal CT in conventional and deep learning-based reconstructions. We found that the addition of a frequency term in the internal noise component of d' is critical to provide a better representation of the radiologists’ performance.

In both FBP and DL reconstruction cases, increasing the lesion size increased the accuracy rate. However, across the 2 image reconstructions for a fixed lesion size, conventional d' did not follow the radiologists’ accuracy rates as closely. We hypothesized that this problem is due to the interconnection of the differing behavior of the NPS of the DL-reconstructed images and internal noise. DL techniques result in a more complex representation of noise, similar to some of the distinct noise patterns in iterative reconstruction algorithms.^{26–28} As shown in this study, with the straight application of standard d' , the

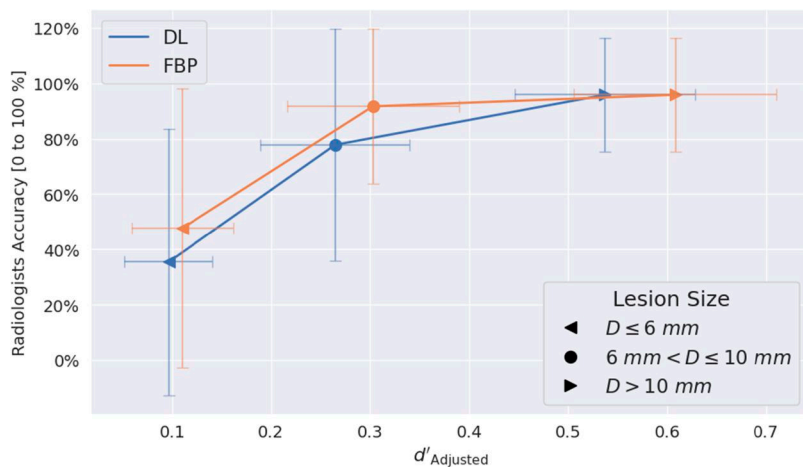


FIGURE 2. Radiologist’s accuracy versus d'_{adj} values. Error bars are the calculated SDs. All the inconsistency issues are addressed with following arrangement of the parameters: $\alpha=0.25$ and $\beta=0.87$ for both FBP and DL-based, and $\gamma=-2.8$ for FBP and $\gamma=-2.4$ for DL-based.

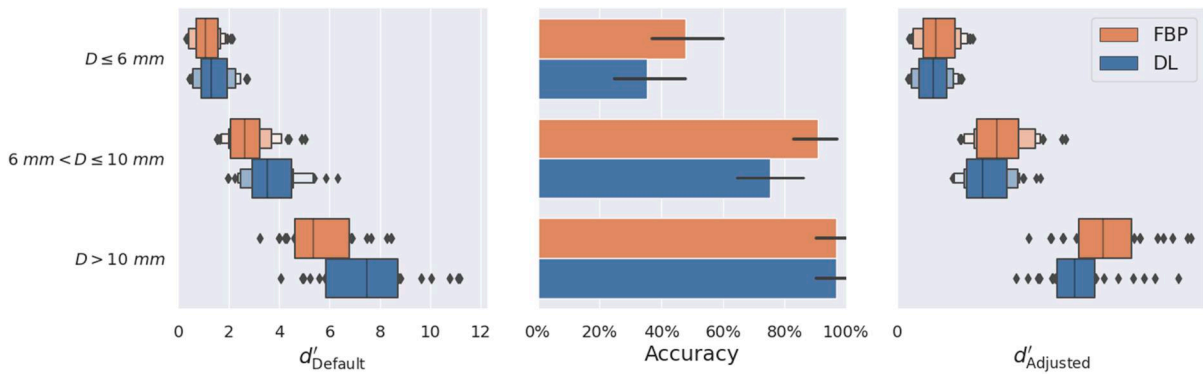


FIGURE 3. Distribution of the d' for the default parameters, radiologist accuracy rate variability, and d'_{adj} distribution across 3 lesion size categories. All the inconsistency issues are addressed with following arrangement of the parameters: $\alpha = 0.25$ and $\beta = 0.87$ for both FBP and DL-based, and $\gamma = -2.8$ for FBP and $\gamma = -2.4$ for DL-based. In the left and right plots, the central box represents the interquartile range (IQR) and the median (vertical line); the whiskers extend to the quartile value within 1.5 times the IQR from the quartiles; and the black dots are individual points outside this range.

discrepancy between the noise patterns can undermine the clinical relevance of d' .

This study proposed a model for the d' anchored and trained upon real radiologist reading of clinical data. Due to the nature of the data, there are limited number of cases and limited number of radiologist readings of the cases, which their inherent statistical fluctuations. Our d' calculations aimed to best represent the data as generated by radiologists. We also additionally investigated the data using observer evaluation for each lesion with respect to characterization and related confidence in the diagnosis analysis keeping track of the individual scores of radiologists. However, we found some discrepancies in the data, including small benign lesions looking more malignant, or small malignant lesions looking more benign. The proposed d'_{adj} essentially aims to account for such inconsistencies. While these discrepancies persist, they still represent the reality of clinical practice, and thus any characterization of radiologist perception in terms of d' needs to include such realities.

Because d' can predict human-observer accuracy, it has assumed a key role in the evaluation of diagnostic imaging.⁸ The recent development and implementation of automated CT image quality measurement techniques have enabled d' calculations in clinical cases *in vivo*. Accuracy of such approaches is pivotal as the results are expected to be used in the optimization of CT procedures. Especially in the implementation of new technologies, the goal is assuring the maintenance of image quality at lowest possible radiation dose. However, inaccurate metrology can undermine that objective. Several studies have shown how some traditional image quality metrics may not always reflect the reader confidence or performance when performing a diagnostic task.¹¹ Therefore, an improved *measured* image quality might not necessarily imply better diagnostic performance. This makes a strong case for careful and anchored formulation of image quality in the context of new technology, in our case, DL reconstruction.

The genetic algorithm (GA) optimization results demonstrated that there is no common hyperparameter

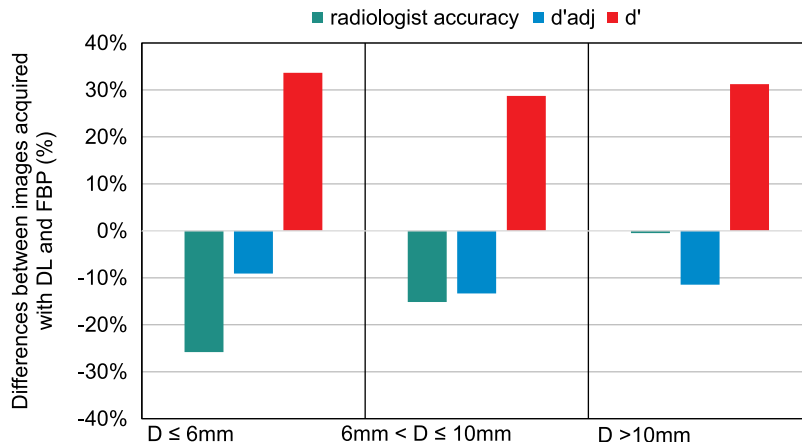


FIGURE 4. Percentage differences between images acquired with DL and images acquired with FBP for radiologists’ accuracy, d'_{adj} , and d' in the 3 lesion groups.

for FBP and DL-based images to address the inconsistency between the radiologist accuracy rate and the d' values. To that extent, Richard and Siewerdsen proposed the intrinsic noise for the FBP kernel to be modeled as we reported in Equation (2) without considering the influence of anatomic background noise.¹⁸ Our study suggested that a new image-dependent parameter γ can be representative of the change in the spectrum of the intrinsic internal noise component. Therefore, γ was embedded into the conventional detectability index. The GA optimization eventually led to the selection of constant α and β values for both FBP and DL-based images, and a variable γ value. This provides a strategy to accommodate the application of observer models to nonlinear image processing methods.

As a limitation, we note that this study included a relatively small sample size of 102 CT studies, which may affect the generalizability of the findings. However, the presented model can be applied to bigger clinical population for a more extensive validation, as well as application to other imaging tasks. In particular, future works, including studies at other dose levels and with additional image reconstruction algorithms or settings, can help determine the range of conditions over which γ remains stable.

CONCLUSION

Detectability index for 2 different CT reconstruction techniques, namely Filtered Back Projection and Deep Learning Image Reconstruction, was calculated applying 2 methodologies and compared with radiologist performance in the assessment of low-contrast liver lesion in abdominal CT. An adjusted d' formalism, including a frequency term for internal noise showed a more consistent association with radiologist accuracy, reflecting the interplay of the NPS, the internal noise, and the eye filter. The new formulation can provide a better representation of the image quality in nonlinear reconstructed CT images.

REFERENCES

- Schwartz FR, Ria F, McCabe C, et al. Image quality of photon counting and energy integrating chest CT - Prospective head-to-head comparison on same patients. *Eur J Radiol*. 2023;166:111014.
- Rajagopal JR, Schwartz FR, McCabe C, et al. Technology characterization through diverse evaluation methodologies: application to thoracic imaging in photon-counting computed tomography. *J Comput Assist Tomogr*. 2024;49:113–124.
- Samei E, Järvinen H, Kortensniemi M, et al. Medical imaging dose optimisation from ground up: expert opinion of an international summit. *J Radiol Prot*. 2018;38:967–989.
- Ria F, Zhang AR, Lerebours R, et al. Optimization of abdominal CT based on a model of total risk minimization by putting radiation risk in perspective with imaging benefit. *Commun Med*. 2024;4:272.
- Christianson O, Winslow J, Frush DP, et al. Automated technique to measure noise in clinical CT examinations. *AJR Am J Roentgenol*. 2015;205:W93–W99.
- Sanders J, Hurwitz L, Samei E. Patient-specific quantification of image quality: an automated method for measuring spatial resolution in clinical CT images. *Med Phys*. 2016;43:5330–5338.
- Abadi E, Sanders J, Samei E. Patient-specific quantification of image quality: an automated technique for measuring the distribution of organ Hounsfield units in clinical chest CT images. *Med Phys*. 2017;44:4736–4746.
- Smith TB, Solomon J, Samei E. Estimating detectability index in vivo: development and validation of an automated methodology. *J Med Imaging*. 2017;5:1–9.
- Cheng Y, Smith TB, Jensen CT, et al. Correlation of algorithmic and visual assessment of lesion detection in clinical images. *Acad Radiol*. 2020;27:847–855.
- Lacy T, Ding A, Minkemeyer V, et al. Patient-based performance assessment for pediatric abdominal CT: an automated monitoring system based on lesion detectability and radiation dose. *Acad Radiol*. 2021;28:217–224.
- Jensen CT, Gupta S, Saleh MM, et al. Reduced-dose deep learning reconstruction for abdominal CT of liver metastases. *Radiology*. 2022;303:90–98.
- Jensen CT, Liu X, Tamm EP, et al. Image quality assessment of abdominal CT by use of new deep learning image reconstruction: initial experience. *AJR Am J Roentgenol*. 2020;215:50–57.
- Jensen CT, Wagner-Bartak NA, Vu LN, et al. Detection of colorectal hepatic metastases is superior at standard radiation dose CT versus reduced dose CT. *Radiology*. 2019;290:400–409.
- Barrett HH, Yao J, Rolland JP, et al. Model observers for assessment of image quality. *Proc Natl Acad Sci U S A*. 1993;90:9758–9765.
- Burgess AE, Xing L, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am A Opt Image Sci Vis*. 1997;14:2420–2442.
- Smith TB, Abadi E, Sauer T, et al. Development and validation of an automated methodology to assess perceptual in vivo noise texture in liver CT. *J Med Imaging*. 2021;8:052113.
- Alsaihati N, Solomon JB, McCrum E, et al. Development, validation, and application of a generic image-based noise addition method for simulating reduced dose computed tomography images. In review. 2025. doi: 10.1002/mp.17444
- Richard S, Siewerdsen JH. Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images. *Med Phys*. 2008;35:5043–5053.
- Zhang Y, Pham BT, Eckstein MP. Evaluation of internal noise methods for Hotelling observer models. *Med Phys*. 2007;34:3312–3322.
- Zhang Y, Leng S, Yu L, et al. Correlation between human and model observer performance for discrimination task in CT. *Phys Med Biol*. 2014;59:3389–3404.
- Yu L, Leng S, Chen L, et al. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms. *Med Phys*. 2013;40:041908.
- Leng S, Yu L, Zhang Y, et al. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. *Med Phys*. 2013;40:081908.
- Solomon J, Lyu P, Marin D, et al. Noise and spatial resolution properties of a commercially available deep learning-based CT reconstruction algorithm. *Med Phys*. 2020;47:3961–3971.
- Fortin FA, De Rainville FM, Garner MA, et al. DEAP: evolutionary algorithms made easy. *J Mach Learn Res*. 2012;13:2171–2175.
- Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput*. 2002;6:182–197.
- Greffier J, Frandon J, Hamard A, et al. Impact of iterative reconstructions on image quality and detectability of focal liver lesions in low-energy monochromatic images. *Phys Med*. 2020;77:36–42.
- Greffier J, Frandon J, Larbi A, et al. Noise assessment across two generations of iterative reconstruction algorithms of three manufacturers using bone reconstruction kernel. *Diagn Interv Imaging*. 2019;100:763–770.
- Greffier J, Larbi A, Frandon J, et al. Comparison of noise-magnitude and noise-texture across two generations of iterative reconstruction algorithms from three manufacturers. *Diagn Interv Imaging*. 2019;100:401–410.