



Estimating subjective evaluation of low-contrast resolution using convolutional neural networks

Yujiro Doi¹ · Atsushi Teramoto² · Ayumi Yamada² · Masanao Kobayashi² · Kuniaki Saito² · Hiroshi Fujita³

Received: 29 May 2021 / Accepted: 4 October 2021 / Published online: 11 October 2021
© Australasian College of Physical Scientists and Engineers in Medicine 2021

Abstract

To develop a convolutional neural network-based method for the subjective evaluation of computed tomography (CT) images having low-contrast resolution due to imaging conditions and nonlinear image processing. Four radiological technologists visually evaluated CT images that were reconstructed using three nonlinear noise reduction processes (AIDR 3D, AIDR 3D Enhanced, AiCE) on a CT system manufactured by CANON. The visual evaluation consisted of two items: low contrast detectability (score: 0–9) and texture pattern (score: 1–5). Four AI models with different convolutional and max pooling layers were constructed and trained on pairs of CANON CT images and average visual assessment scores of four radiological technologists. CANON CT images not used for training were used to evaluate prediction performance. In addition, CT images scanned with a SIEMENS CT system were input to each AI model for external validation. The mean absolute error and correlation coefficients were used as evaluation metrics. Our proposed AI model can evaluate low-contrast detectability and texture patterns with high accuracy, which varies with the dose administered and the nonlinear noise reduction process. The proposed AI model is also expected to be suitable for upcoming reconstruction algorithms that will be released in the future.

Keywords Image quality · Computed tomography · Deep learning · Convolutional neural network · Subjective evaluation

Introduction

In recent years, tube current control technology, image reconstruction technology, and image filtering technology have made it possible to significantly reduce the amount of radiation humans are exposed to during computed tomography (CT) scans. In particular, because noise reduction processing based on iterative reconstruction (IR) algorithms can reduce radiation exposure while maintaining image quality [1], different algorithms are being developed by manufacturers [2, 3]. The effect of IR algorithms on image quality is known to be nonlinear and to produce a different frequency

response than conventional filtered back projection (FBP) reconstruction [2, 4]. In addition, the emergence of image processing using artificial intelligence is expected to accelerate the non-linearization of images.

Objective evaluation methods for low contrast detectability include image noise and the contrast-to-noise ratio (CNR). However, it has been reported that the change in low-contrast detectability due to IR does not correlate with image noise or CNR [5–8], suggesting that it is difficult to evaluate IR based on physical assessment alone. Furthermore, it has been reported that the noise reduction process applying the IR method changes the texture pattern of the image at low doses, resulting in a visually different impression compared to FBP [9]. For this reason, many phantom studies that focus on the evaluation of IR combine objective and subjective evaluations [10].

For the same reason, it is necessary to determine the scan parameters based on a comprehensive judgment based on both objective and subjective evaluation results in clinical practice. However, subjective evaluation is burdensome because the observer has to evaluate a large number of images one by one [5, 10]. Furthermore, because the characteristics of image quality differ among CT manufacturers,

✉ Atsushi Teramoto
teramoto@fujita-hu.ac.jp

¹ Department of Radiology, Fujita Health University Hospital, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake City, Aichi 470-1192, Japan

² Graduate School of Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake City, Aichi 470-1192, Japan

³ Faculty of Engineering, Gifu University, 1-1 Yanagido, Gifu-City, Gifu 501-1194, Japan

subjective evaluations must be conducted for each CT system. For these reasons, there is a need to establish a technology that provides automated subjective evaluation using certain criteria.

Convolutional neural networks (CNNs), which are an artificial intelligence technology that simulates the functions of human neurons and human vision, achieve high performance in image recognition and are widely used in the field of image recognition processing [11, 12]. Many studies have been published on image quality evaluation using neural networks [13, 14], and some have been applied to medical imaging [15, 16]. Kim et.al reported as detecting low contrast object using CNN in phantom study [17]. However, Kim's study used CT images obtained from a single CT system, and it is not specified whether the results can be obtained with high accuracy for reconstruction algorithms from different CT manufacturers and for unknown image reconstructions. In addition, the cropping process for each object is very complicated, and it is necessary to develop a simple evaluation system that can input images without preprocessing.

In this study, we developed a CNN-based method for the subjective evaluation of CT images to detect changes in low contrast resolution due to scan parameters and nonlinear image processing.

Method

Phantom

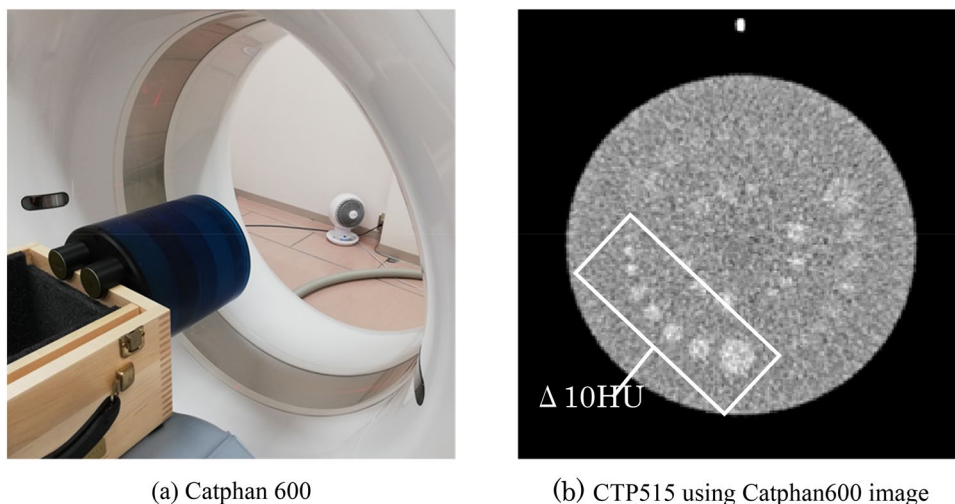
We used the Catphan600 (The Phantom Laboratory, Salem, NY, USA) consisting of five modules (CTP404 Slice Geometry and Sensitometry Module, CTP591 Bead Geometry Module, CTP528 High Resolution Module, CTP515 Low Contrast Module, and CTP486 Uniformity Module), which

enable different measurements depending on the imaging cross section. The CTP515 module used in this study contains cylindrical rods (z-length=40 mm) with diameters of 2–15 mm and contrasts of 3HU, 5HU, and 10HU (Fig. 1). The observer can evaluate the low contrast resolution by visually examining a CT image of a phantom in which a number of these cylindrical rods having different diameters and contrasts are embedded. In this study, cylindrical rods with a contrast difference of 10HU were used for observer evaluation.

CT systems and data acquisition parameters

This study was performed using three different CT systems from two different manufacturers. First, we used the Aquilion ONE Genesis system (Canon Medical Systems). Acquisition was performed using a 120 kV helical scan (beam pitch: 0.637). The tube current was set to 640 mA when the rotation time was 1.5 s/rot and was set to 650, 490, 340, 250, 130, 60, 60, 30, 15, or 10 mA when the rotation time was 1.0 s/rot. The beam collimation was set as 80×0.5 mm, slice thickness was 0.5 mm, the field of view (FOV) was 240 mm, and the matrix size was 512×512 pixels. Images were reconstructed using FBP, adaptive iterative dose reduction 3D (AIDR 3D and AIDR 3D Enhanced), and the advanced intelligent clear-IQ engine (AiCE). The kernels for FBP, AIDR 3D, and AIDR3D Enhanced were all FC03. AIDR 3D and AIDR 3D Enhanced, which are noise reduction processes using the IR method, allow multiple noise reduction levels, which are set as follows: AIDR 3D (weak, mild, standard, or strong) and AIDR 3D enhanced (e-mild, e-standard, or e-strong). AiCE, which is a noise reduction technology using deep learning, can provide different noise reduction intensities depending on the body part, and was set to body-mild, body-standard, or body-strong.

Fig. 1 Catphan600 and CT image of CTP515



Second, we used the SOMATOM Force system (SIEMENS). Acquisition was performed using a 120 kV helical scan (beam pitch: 0.6). The tube current was set to 630, 420, 315, 209, 157, 79, or 41 mA when the rotation time was 1.0 s/rot, and to 38 mA when the rotation time was 0.5 s/rot. We set the beam collimation as 192×0.6 mm, the slice thickness as 0.6 mm, the FOV as 240 mm, and the matrix size as 512×512 . Images were reconstructed using FBP and advanced modeled iterative reconstruction (ADMIRE). ADMIRE, a noise reduction process using the IR method, can be set to several noise reduction levels, and was set to Strength 1, 2, 3, 4, or 5. The kernel used was Br32.

Finally, the SOMATOM AS system (SIEMENS) was used, and acquisition was performed using a 120 kV

helical scan (beam pitch: 0.6). The tube current was 626, 416, 312, 207, 156, 78, or 38 mA at 1.0 s/rot and 38 mA at 0.5 s/rot. We set the beam collimation as 128×0.6 mm, the slice thickness as 0.6 mm, the FOV as 240 mm, and the matrix size as 512×512 . Similar to ADMIRE described above, SAFIRE was used for reconstruction and can be set to multiple noise reduction levels, which were set to Strength 1, 2, 3, 4, or 5. For the kernels, H30 medium smooth was used for FBP, and J30 medium smooth was used for SAFIRE. Table 1 lists the scan parameters described above.

Table 1 Scanning and reconstruction parameters

Vendor	CANON	SIEMENS	
	Aquilion ONE Genesis	SOMATOM FORCE	SOMATOM AS
Tube voltage (kV)	120	120	120
Tube	Tube current (mAs)/CTDI (mGy)	Tube current (mAs)/CTDI (mGy)	Tube current (mAs)/CTDI (mGy)
No. 1	640/89.2	630/89.94	626/89.97
No. 2	650/60.4	420/59.96	416/59.98
No. 3	490/45.5	315/44.97	312/45.11
No. 4	340/30.2	209/29.98	207/29.99
No. 5	250/22.2	157/22.61	156/22.55
No. 6	130/11.6	79/11.42	78/11.28
No. 7	60/5.3	41/5.59	38/5.52
No. 8	30/2.7	19/2.74	19/2.76
No. 9	15/1.3	–	–
No. 10	10/0.9	–	–
No. 11	10/0.3	–	–
Rotation time (s)	1.5 (No. 1) 1.0 (No. 2–10) 0.35 (No. 11)	1.0 (No. 1–7) 0.5 (No. 8) –	1.0 (No. 1–7) 0.5 (No. 8) –
Reconstruction	FBP	FBP	FBP
IR 1	AIDR 3D (weak, mild, standard, strong)	ADMIRE (Strength 1, 2, 3, 4, 5)	SAFIRE (Strength 1, 2, 3, 4, 5)
IR 2	AIDR 3D enhanced (e-mild, e-standard, e-strong)	–	–
DLR	AiCE (Body-mild, body-standard, body-strong)	–	–
MBIR	FIRST (Body-mild, body-standard, body-strong)	–	–
Collimation	0.5 mm \times 80 rows	0.6 mm \times 192 rows	0.6 mm \times 128 rows
Pitch	0.625	0.6	0.6
Display FOV (mm)	240	240	240
Reconstruction filters	FC 13 (FBP, IR1, IR2) None of DLR and MBIR	Br32 (FBP, IR1) –	H30 medium smooth (FBP) J30 medium smooth (IR1)
Thickness (mm)	0.5	0.6	0.6
Gap (mm)	0.5	0.5	0.5

Labeling: human observer tests

Visual evaluation of CTP515 images was performed by four observers. Images were taken from slices where the markers on the top of the phantom were visible. Visual evaluation was performed on two items: low contrast detectability and texture pattern.

For low contrast detectability, the number of visible rods was recorded based on a total of nine rods having different diameters, all rods having a 10 HU ($\Delta 10\text{HU}$) difference from the background CT value (rod diameters: 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, and 15.0 mm) (Fig. 2). Herein, “visible” indicates that the rod was recognized by observers irrespective of the shape.

Singh et al. reported that blotchy pixilated artifacts produced by the IR method contribute to an observer's visual discomfort [9, 18]. It is likely that this change in texture pattern is not limited to the IR method, but can also be caused by other nonlinear noise reduction processes. Therefore, we examined the changes in texture patterns caused by various nonlinear noise reduction processes. The degree of blotchiness and artificial texture of each cylindrical rod ($\Delta 10\text{HU}$, 15 mm ϕ) and its surroundings was evaluated on a 5-point scale (Table 2, Fig. 3). However, if the rod was not visible because of high image noise, the score was set to 1. The observer experiment was conducted by four radiological technologists with 4–15 years of experience. The images were displayed using a window level of 50 and window width of 50. The viewing distance was not specified, and the evaluation was conducted at a distance that was comfortable for each observer.

Table 2 Evaluation criteria for texture pattern

Score	Rod visibility	Blotchiness/artificial texture
1	No	–
2	Yes	Very strong
3	Yes	Strong
4	Yes	Weak
5	Yes	None or very weak

The evaluation was conducted on a tablet device (Surface GO, Microsoft) equipped with the software developed for image observations. An example of the screen produced by the application software is shown in Fig. 4. Images were randomly displayed on the left side of Fig. 4, and the evaluation score was assigned by the observer using a graphical user interface on the right side. The score of an image was the average of four observer's scores.

Network architecture

For the subjective evaluation, we adopted an original CNN model that regresses the scores obtained for low contrast detectability and texture patterns in the output layer with continuous values from 0 to 1 [12]. Both the window width (WW) and window level (WL) of the given CT images were set to 50 and converted to 8-bit grayscale images.

A CNN includes input layers, hidden layers (convolutional layer, max pooling layer, and fully connected layer), and output layers. The hidden layer is responsible for extracting features, and the features increase in the number of layers within the hidden layer. However,

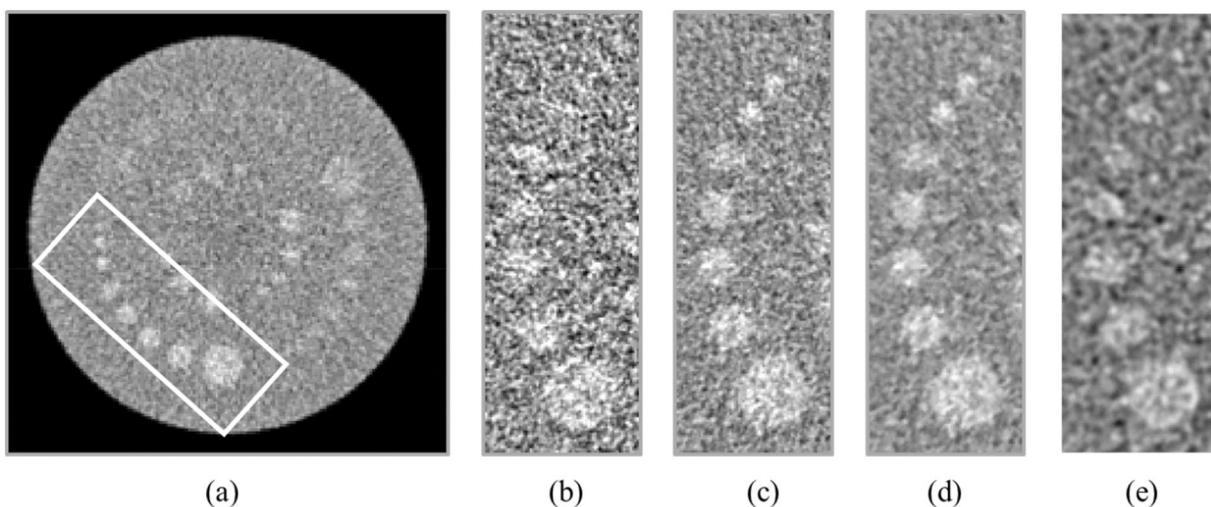


Fig. 2 Evaluation of low contrast detectability and image examples using a phantom. **a** Overall view of CTP515 image, **b** $\Delta 10\text{HU}$ cylindrical rods (CANON, 250 mAs, FBP), **c** $\Delta 10\text{HU}$ cylindrical rods

(CANON, 650 mAs, FBP), **d** $\Delta 10\text{HU}$ cylindrical rods (CANON, 650 mAs, AIDR 3D: strong setting), **e** $\Delta 10\text{HU}$ cylindrical rods (SIEMENS, 420 mAs, ADMIRE)

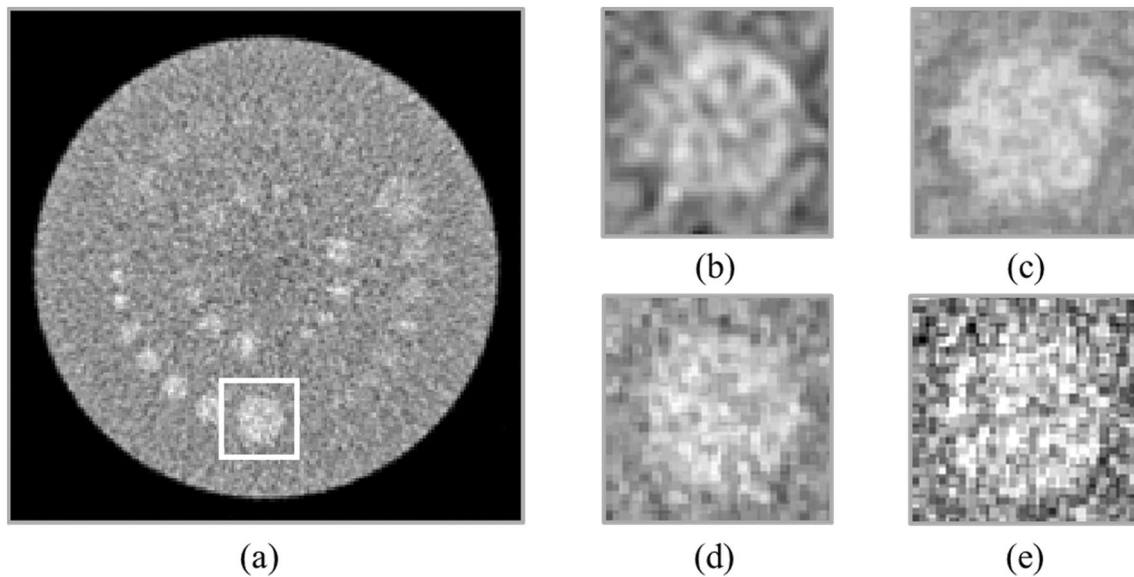


Fig. 3 Evaluation of texture patterns. **a** Overall view of CTP515 image, **b** Cylindrical rod with $\Delta 10\text{HU}$ (CANON, 60.4 mAs, FIRST), **c** cylindrical rod with $\Delta 10\text{HU}$ (CANON, 960 mAs, AiCE), **d** cylin-

drical rod with $\Delta 10\text{HU}$ (CANON, 650 mAs, AIDR 3D: strong setting), **e** cylindrical rod with $\Delta 10\text{HU}$ (CANON, 250 mAs, FBP)

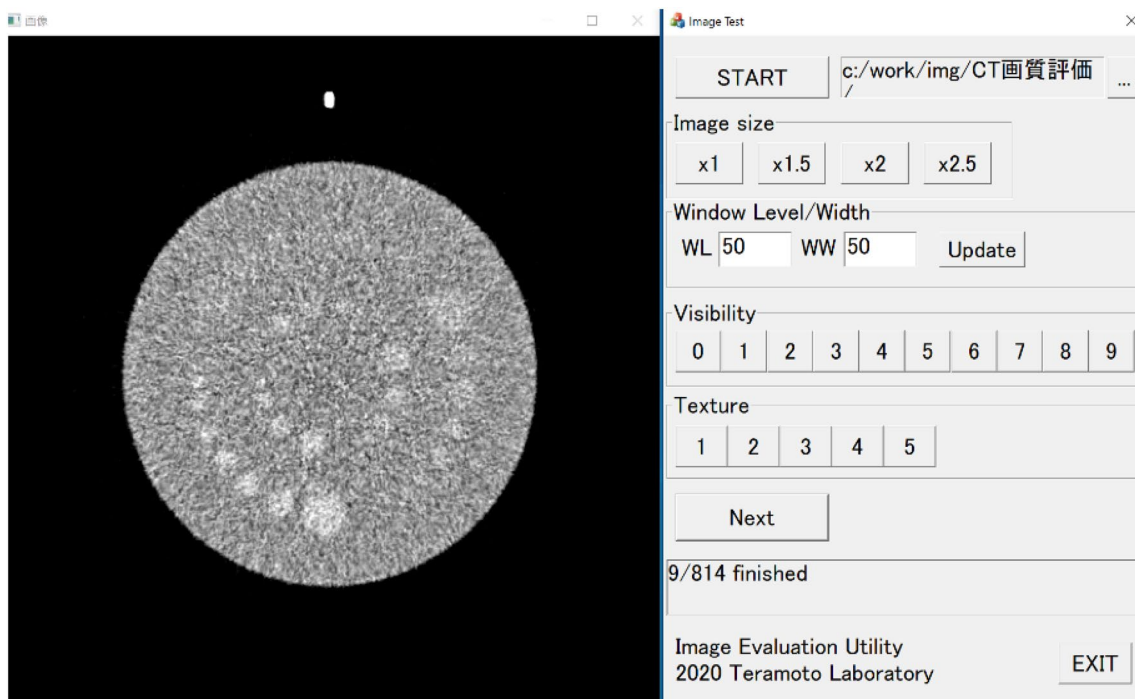


Fig. 4 Example screen using application software for image quality evaluation

excessive extraction features result in overfitting and affect the prediction performance of the CNN. Therefore, to optimize the evaluation of low contrast detectability and texture, we compared four CNN models with different hidden layers. As a baseline, we created AI model 1,

a CNN model with three convolutional and pooling layers and two fully connected layers. Furthermore, we created AI model 2 with four convolutional and pooling layers, AI model 3 with five convolutional and pooling layers, and AI model 4 with six convolutional and pooling layers, and

compared their prediction performance (Fig. 5). In the convolutional layer, we set the filter size to 3×3 , stride of 1, and max pooling matrix size of 2×2 . The activation function was the rectified linear unit (ReLU) function.

Training

The CT images used for training were cropped to 350×350 pixels to remove the background around the phantom. Out of a total of 407 images of FBP, AIDR 3D, AIDR 3D Enhanced, and AiCE reconstructed on Aquilion ONE, 308 images were used for training each model and 99 were used for validation. The cross-entropy error was used as the loss

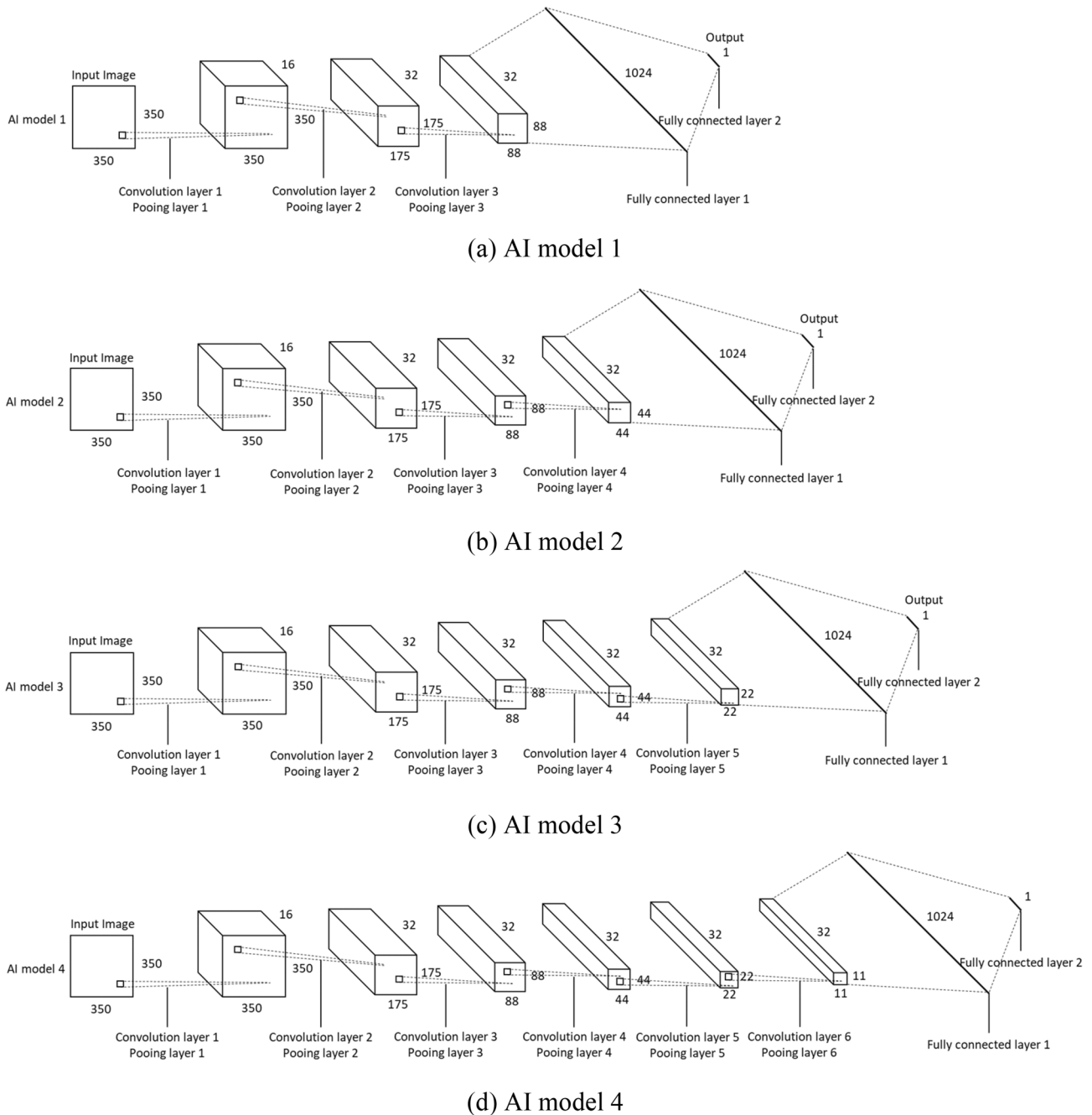


Fig. 5 CNN architectures for image assessment. **a** Convolutional layers: 3, pooling layers: 3, fully connected layers: 2. **b** Convolutional layers: 4, pooling layers: 4, fully connected layers: 2. **c** Convolutional

layers: 5, pooling layers: 5, fully connected layers: 2. **d** Convolutional layers: 6, pooling layers: 6, fully connected layers: 2

function, and Adam was used as the optimization algorithm. The number of training epochs and learning coefficient were set to 50 and 0.001, respectively. The training process was performed using TensorFlow [19] and Keras frameworks [20] running in a graphics processing unit (GPU) environment. The computers used for training were equipped with an AMD Ryzen 9 3950X processor and an NVIDIA TITAN RTX GPU.

Contrast-to-noise ratio

We measured CNR for 11 computed tomography dose index (CTDI) and four reconstructions (FBP, AIDR 3D, AIDR 3D Enhanced, and AiCE) in CANON CT images. CNR is expressed as

$$\text{CNR} = \frac{CT_{Rod} - CT_{BG}}{SD_{BG}}$$

where CT_{Rod} denotes the CT value of the rod ($\Delta 10\text{HU}$, 15 mm ϕ) used for the human observer test, CT_{BG} denotes the CT value of the background near the rod, and SD_{BG} denotes the standard deviation of the background near the rod. The ROI was set to 20×20 . The CNR score was the average CNR of the three CT images. For low contrast detectability, we calculated correlation diagrams between the CNR and the score of the human observer test.

Evaluation metrics

To evaluate the prediction performance of each model, 99 images (not used for training) out of a total of 407 images produced using FBP, AIDR 3D, AIDR 3D Enhanced, and AiCE reconstructed by Aquilion ONE Genesis were used as input for each model, and the correlation diagrams of the scores between the observer and the CNN outputs were obtained. To conduct the external validation, 288 CT images acquired by SIEMENS scanners (ADMIRE: 144 images, SAFIRE: 144 images), which were not used for training, were input into each CNN model and the correlations of scores between the observer and the CNN outputs were obtained. The mean absolute error (MAE) and correlation coefficient were then calculated from the results. MAE is an index used to measure the accuracy of a prediction model and is expressed as

$$\text{MAE} = \frac{1}{n} \sum_{k=1}^n (f_i - y_i)^2$$

where f_i is the actual value, y_i is the predicted value, and n is the number of samples. The correlation coefficient is an index that measures the linear relationship between two variables and is expressed as

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2} \sqrt{\sum_{k=1}^n (y_k - \bar{y})^2}}$$

where r is the correlation coefficient, x and y are target variables, and i, j , and k are the index values related to the sample number.

Results

Contrast-to-noise ratio

The CNR for each CTDI was increased by the nonlinear noise reduction process, as shown in Fig. 6. Based on the correlation between the CNR and the score of human observers, we observed that the correlation coefficient was different for each CTDI (Fig. 7 and Table 3). For all CTDI, the correlation coefficient between the CNR and the score of human observers was 0.92 (Table 3).

Predictive performance of the models

We evaluated 99 CT images with AI models 1 to 4, which were trained to output the evaluation values of low contrast detectability and texture pattern, and then calculated the correlations with the subjective evaluation results of the observers. The results are shown in Fig. 8, and the MAEs and correlation coefficients are listed in Table 4. In the plotted points of Fig. 8, the x coordinates of certain points have non-integer values because we plotted points using average scores with four observers. AI model 2 achieved the best results in terms of low contrast detectability, with a MAE of 0.458 and a correlation coefficient of 0.967. With respect to the texture patterns, AI model 3 achieved the best results, with a MAE of 0.227 and a correlation coefficient of 0.972. Examples of CT images and the scores as evaluated by the observer and the best model are shown in Fig. 9.

External validation

We input 288 CT images reconstructed by SAFIRE and ADMIRE into each AI model and calculated the correlation diagram with the subjective evaluation results of the observer (Fig. 10). AI model 2 achieved the best results in terms of low contrast detectability, with a MAE of 0.700 and a correlation coefficient of 0.932. For texture patterns, AI model 4 achieved the best overall results, with a MAE of 0.493 and a correlation coefficient of 0.866.

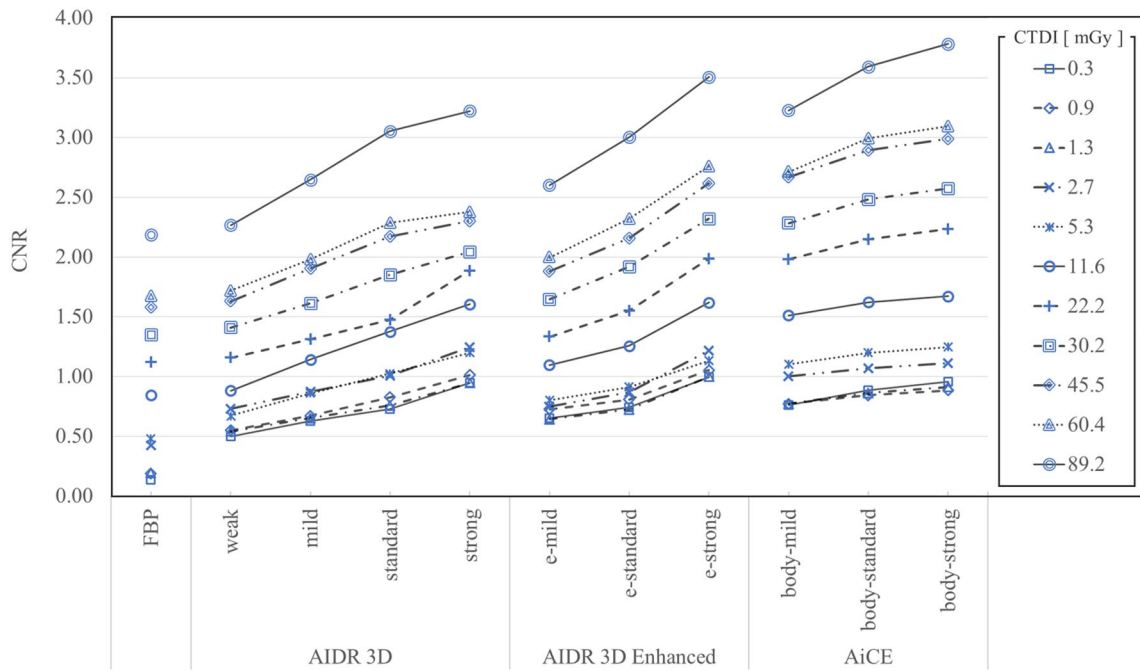
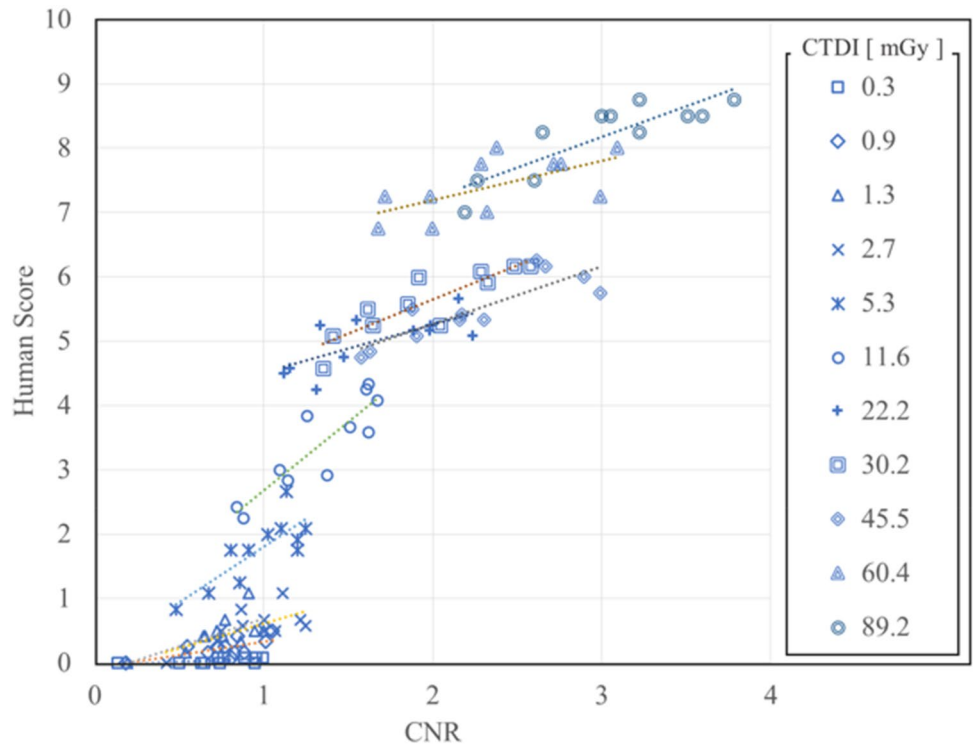


Fig. 6 Contrast-to-noise ratio between CTDI and reconstructions

Fig. 7 Correlation for each CTDI between CNR and observer’s visual evaluations. The approximate line for each CTDI was added as an eye guide



Discussion

Nonlinear noise reduction processes are known to change the resolution and noise reduction performance depending

on the dose level used and on the contrast of a CT image [4]. Image noise and resolution are important parameters that determine image quality and have a significant impact on the detectability of lesions and the depiction of structures under low contrast conditions [21]. In addition,

Table 3 Correlation coefficient for each CTDI

CTDI (mGy)	Correlation coefficient
0.3	0.56
0.9	0.69
1.3	0.70
2.7	0.68
5.3	0.82
11.6	0.89
22.2	0.71
30.2	0.87
45.5	0.86
60.4	0.64
89.2	0.86
All	0.92

In this study, we proposed a technique to evaluate the performance of CNN on image observation of processed CT images based on certain criteria using linear FBP and non-linear noise reduction techniques by comparing the CNN performance to expert human observers.

First, we evaluated the relationship between the CNR and the human observers using CANON's CT images. The correlation between the CNR and the subjective evaluation of observers differed according to dose. As for the relationship between the CNR and the visual evaluation at low doses, when rods could not be detected under FBP conditions, the detectability of rods was hardly improved by nonlinear noise reduction processing. However, the CNR improved with increasing strength of the nonlinear noise reduction process at all doses and reconstructions. In summary, there is a non-linear relationship between CNR

because the nonlinear noise reduction process has multiple parameters related to the achievable degree of noise [3], it is difficult for an observer to evaluate all the parameters based on fixed criteria. In addition to nonlinear noise reduction processing, various applications and imaging technologies, such as automatic exposure control technology [22], high pitch helical scanning [23], and low-tube voltage imaging in contrast-enhanced CT [24], as well as proposed new technologies are expected to preserve image quality while reducing radiation exposure in the future. Against this background, we consider it important to establish a uniform evaluation method for unknown technologies that are expected to be developed in the future.

Table 4 Mean absolute error and correlation coefficient—measuring accuracy of AI model for known reconstruction

AI model	Low contrast detectability		Texture pattern	
	Mean absolute error	Correlation coefficient	Mean absolute error	Correlation coefficient
Model 1	0.679	0.948	0.530	0.882
Model 2	0.458	0.968	0.329	0.939
Model 3	0.547	0.961	0.227	0.973
Model 4	1.482	0.877	0.254	0.967

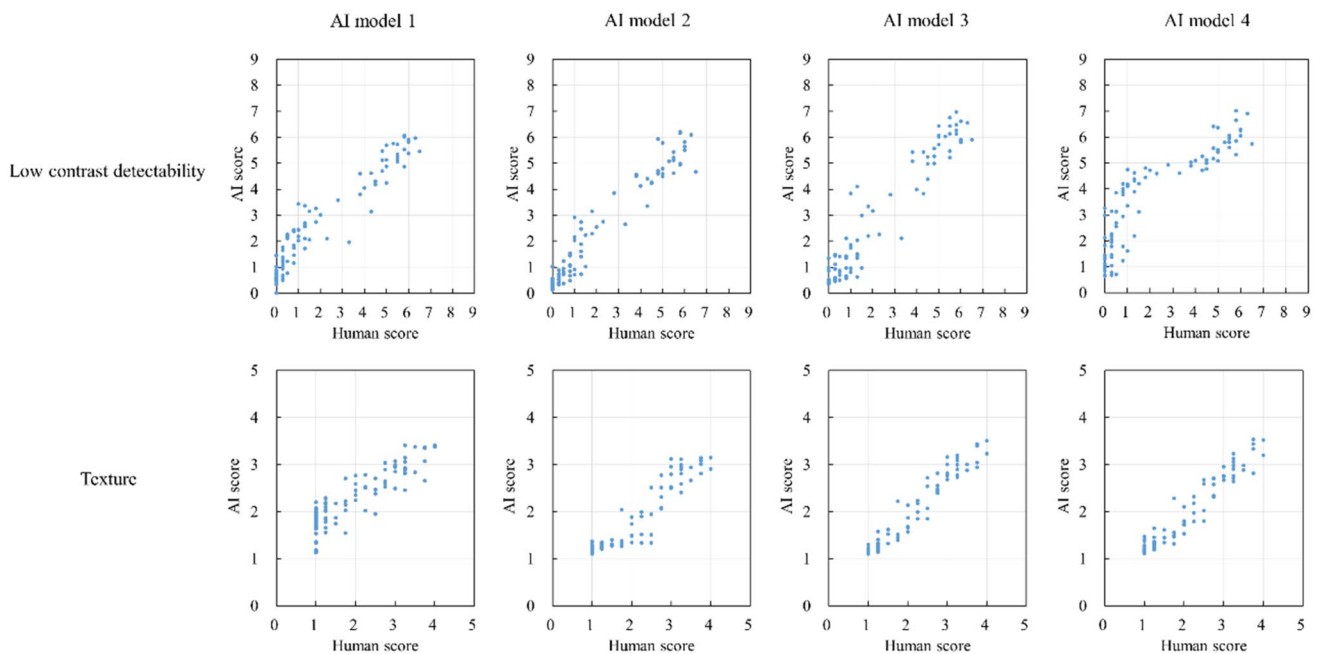
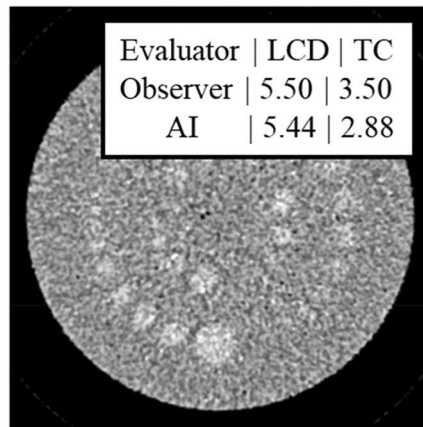


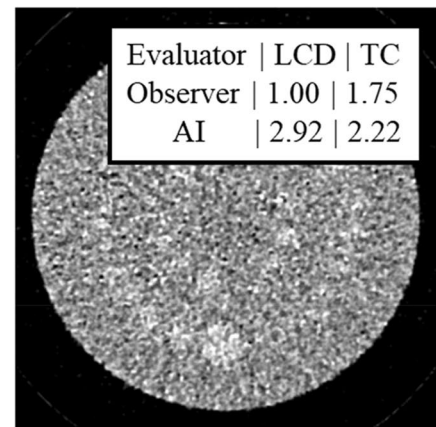
Fig. 8 Correlation of visual evaluation scores between observers and AI—measuring accuracy of AI model for known reconstruction process

Fig. 9 Examples of CT images and scores for low contrast detectability and texture. **a** Canon/250 mAs/AiCE_Mild. **b** Canon/60 mAs/AiCE_Strong. **c** Canon/60 mAs/AIDR 3D Enhanced_Strong. **d** Canon/130 mAs/AIDR 3D_Standard

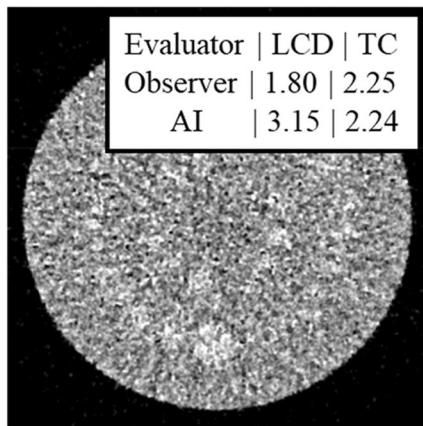
LCD : Low Contrast Detectability TC : Texture



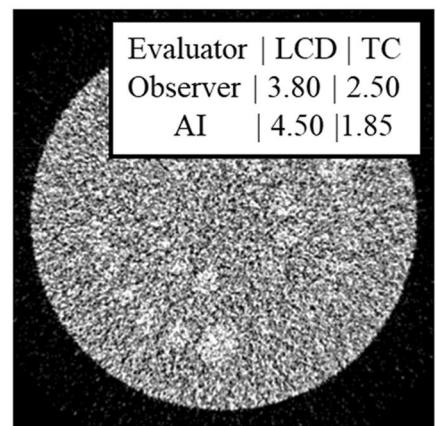
(a)



(b)



(c)



(d)

results and subjective evaluation results; we believe that it is difficult to evaluate the human observer test using CNR. AI model 2, which was trained using CANON CT images, showed a correlation coefficient of 0.968, as presented in Table 4, and was determined to be appropriate for evaluation using a CNN instead of CNR.

When evaluating low contrast detectability, the minimum MAE between the subjective evaluation by the observer and the prediction result by a CNN model was 0.458 and the correlation coefficient was 0.967, indicating that low contrast detectability can be accurately evaluated by the best CNN model (AI model 2) among the four models evaluated. Noise and contrast are relatively simple image quality characteristics, and we believe that the feature extraction and feature integration processes inherent in CNN worked well in this case. Because favorable results were obtained even when the images were input directly into a CNN without any image processing such as cropping or edge detection, we believe that this method is a

versatile tool that can be applied to other image quality evaluation properties.

The AI model trained on CANON's CT images was validated using CT images obtained from SIEMENS' scanners to determine whether it would perform satisfactorily on unknown reconstructed images from CANON and other manufactured systems that will be released in the future. The subjective evaluation by the observers and the prediction results using CNN showed good results in both MAE and correlation coefficient, so we believe that this AI model is suitable for use in upcoming image reconstruction techniques. To optimize the AI model with low contrast resolution, we evaluated four AI models with different network architectures. In this study, we performed the internal validation of the prediction performance of the AI model using CANON CT images (Table 4) and the external validation of the generalization performance of the AI model using SIEMENS CT images (Table 5). Because the MAE and correlation coefficient showed the best value (Tables 4, 5), AI

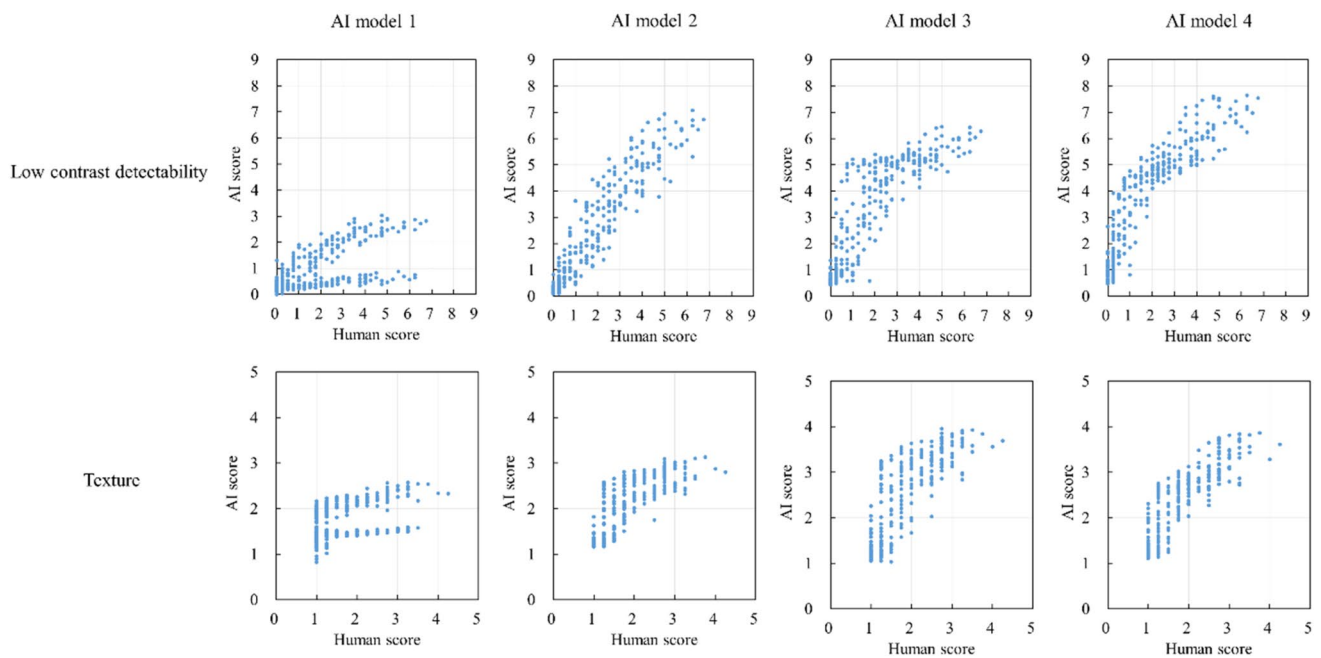


Fig. 10 Correlation of visual evaluation between observers and AI model—measuring accuracy of AI model for unknown reconstruction process

Table 5 Mean absolute error and correlation coefficient—measuring accuracy of AI model for unknown reconstruction

AI model	Low contrast detectability		Texture pattern	
	Mean absolute error	Correlation coefficient	Mean absolute error	Correlation coefficient
Model 1	1.138	0.613	0.545	0.451
Model 2	0.700	0.932	0.351	0.837
Model 3	1.338	0.842	0.585	0.829
Model 4	1.949	0.879	0.493	0.866

model 2 was considered as the most optimized model in terms of low contrast detectability. In terms of texture pattern, AI Model 3 and AI Model 4 obtained satisfactory values, as presented in Table 4, and AI Model 2 and AI Model 4 obtained acceptable values, as presented in Table 5, suggesting that AI Model 4 was the best model for overall judgment.

This study has some limitations. First, the results of the visual evaluations by the observers may contain bias owing to the geometric arrangement of the phantoms. Because the main purpose of this study was to confirm the correlation between the score of an observer and the visual evaluation by CNN, bias resulting from the geometric arrangement and a detailed study of the visual evaluation process were excluded from this study. Moreover, bias may have occurred because of a small number of observers. Therefore, we have the following challenges to obtain more precise results: (1) increasing the number of observers, (2) introduction of samples where the presence of the signal is not known to

the observer, and (3) changing the evaluation method from evaluating the number of visible rods to evaluating the visibility of each rod and performing ROC analysis.

Conclusion

In conclusion, we have proposed a method for the automatic visual evaluation of low contrast resolution images using an AI model based on CNN. Proposed AI models can accurately evaluate the low contrast detectability and texture pattern in an image, which vary with dose level and the nonlinear noise reduction process applied. Proposed AI models are also expected to be suitable for any upcoming reconstruction algorithms that may be released in the future.

Acknowledgements We would like to thank Mr. Matsumoto for useful discussions. We would like to express the deepest appreciation to Mr. Mori and Mr. Yamamoto, department of radiology, Toyama university hospital. We would like to thank Editage (www.editage.com) for English language editing.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

Code availability The codes we write during the current study available from the corresponding author on reasonable request.

Declarations

Conflict of interest This study was supported in part by the research grant from A-Line corporation, Osaka, Japan.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate This article does not contain patient data.

Consent for publication Not applicable.

References

- Singh S, Kalra MK, Hsieh J, Licato PE, Do S, Pien HH, Blake MA (2010) Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. *Radiology* 257(2):373–383. <https://doi.org/10.1148/radiol.10092212>
- Löve A, Olsson ML, Siemund R, Stålhammar F, Björkman-Burtscher IM, Söderberg M (2013) Six iterative reconstruction algorithms in brain CT: a phantom study on image quality at different radiation dose levels. *Br J Radiol* 86(1031):388. <https://doi.org/10.1259/bjr.20130388>
- Willeminck MJ, de Jong PA, Leiner T, de Heer LM, Nievelstein RA, Budde RP, Schilham AM (2013) Iterative reconstruction techniques for computed tomography. Part 1: technical principles. *Eur Radiol* 23(6):1623–1631. <https://doi.org/10.1007/s00330-012-2765-y>
- 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(7):4115–4122. <https://doi.org/10.1118/1.4725171>
- Schindera ST, Odedra D, Raza SA, Kim TK, Jang HJ, Szucs-Farkas Z, Rogalla P (2013) Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology* 269(2):511–518. <https://doi.org/10.1148/radiol.13122349>
- Schindera ST, Odedra D, Mercer D, Thipphavong S, Chou P, Szucs-Farkas Z, Rogalla P (2014) Hybrid iterative reconstruction technique for abdominal CT protocols in obese patients: assessment of image quality, radiation dose, and low-contrast detectability in a phantom. *Am J Roentgenol* 202(2):W146–152. <https://doi.org/10.2214/AJR.12.10513>
- Baker ME, Dong F, Primak A, Obuchowski NA, Einstein D, Gandhi N, Herts BR, Purysko A, Remer E, Vachhani N (2012) Contrast-to-noise ratio and low-contrast object resolution on full- and low-dose MDCT: SAFIRE versus filtered back projection in a low-contrast object phantom and in the liver. *AJR* 199(1):8–18. <https://doi.org/10.2214/AJR.11.7421>
- Kondo M, Hatakenaka M, Higuchi K, Fujioka T, Shirasaka T, Nakamura Y, Nakamura K, Yoshiura T, Honda H (2013) Feasibility of low-radiation-dose CT for abdominal examinations with hybrid iterative reconstruction algorithm: low-contrast phantom study. *Radiol Phys Technol* 6(2):287–292. <https://doi.org/10.1007/s12194-012-0197-7>
- Yang WJ, Yan FH, Liu B, Pang LF, Hou L, Zhang H, Pan ZL, Chen KM (2013) Can sinogram-affirmed iterative (SAFIRE) reconstruction improve imaging quality on low-dose lung CT screening compared with traditional filtered back projection (FBP) reconstruction? *J Comput Assist Tomogr* 37(2):301–305. <https://doi.org/10.1097/RCT.0b013e31827b8c66>
- Mileto A, Zamora DA, Alessio AM, Pereira C, Liu J, Bhargava P, Carnell J, Cowan SM, Dighe MK, Gunn ML, Kim S, Kolokythas O, Lee JH, Maki JH, Moshiri M, Nasrullah A, O'Malley RB, Schmiedl UP, Soloff EV, Toia GV, Wang CL, Kanal KM (2018) CT Detectability of small low-contrast hypoattenuating focal lesions: iterative reconstructions versus filtered back projection. *Radiology* 289(2):443–454. <https://doi.org/10.1148/radiol.2018180137>
- LeCun Y, Bengio Y, Hinton GE (2015) Deep learning. *Nature* 512:436–444
- Matsubara N, Teramoto A, Saito K, Fujita H (2020) Bone suppression for chest X-ray image using a convolutional neural filter. *Phys Eng Sci Med* 43:97–108. <https://doi.org/10.1007/s13246-019-00822-w>
- Kim J and Lee S (2017) Deep learning of human visual sensitivity in image quality assessment framework. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, pp 1969–1977. <https://doi.org/10.1109/CVPR.2017.213>
- Hou W, Gao X, Tao D, Li X (2015) Blind image quality assessment via deep learning. *IEEE Trans Neural Netw Learn Syst* 26(6):1275–1286. <https://doi.org/10.1109/TNNLS.2014.2336852>
- Lauermann JL, Treder M, Alnawaiseh M, Clemens CR, Eter N, Alten F (2019) Automated OCT angiography image quality assessment using a deep learning algorithm. *Graefes Arch Clin Exp Ophthalmol* 257(8):1641–1648. <https://doi.org/10.1007/s00417-019-04338-7>
- Sujit SJ, Coronado I, Kamali A, Narayana PA, Gabr RE (2019) Automated image quality evaluation of structural brain MRI using an ensemble of deep learning networks. *J Magn Reson Imaging* 50(4):1260–1267. <https://doi.org/10.1002/jmri.26693>
- Kim HY, Lee K, Chang W, Kim Y, Lee S, Oh DY, Sunwoo L, Lee YJ, Kim YH (2021) Robustness of deep learning algorithm to varying imaging conditions in detecting low contrast objects in computed tomography phantom images: in comparison to 12 radiologists. *Diagnostics (Basel, Switzerland)* 11(3):410. <https://doi.org/10.3390/diagnostics11030410>
- Singh S, Kalra MK, Gilman MD, Hsieh J, Pien HH, Digumarthy SR, Shepard JA (2011) Adaptive statistical iterative reconstruction technique for radiation dose reduction in chest CT: a pilot study. *Radiology* 259(2):565–573. <https://doi.org/10.1148/radiol.11101450>
- Abadi M, Barham P, Chen J et al (2016) Tensorflow: a system for large-scale machine learning. *OSDI* 16:265–283
- Chollet F et al (2018) Keras: the python deep learning library. *Astrophysics Source Code Libr*.
- Samei E, Bakalyar D, Boedeker KL, Brady S, Fan J, Leng S, Myers KJ, Popescu LM, Ramirez Giraldo JC, Ranallo F, Solomon J, Vaishnav J, Wang J (2019) Performance evaluation of computed tomography systems: summary of AAPM Task Group 233. *Med Phys* 46(11):735–756. <https://doi.org/10.1002/mp.13763>
- Rizzo S, Kalra M, Schmidt B, Dalal T, Suess C, Flohr T, Blake M, Saini S (2006) Comparison of angular and combined automatic tube current modulation techniques with constant tube current CT of the abdomen and pelvis. *Am J Roentgenol* 186(3):673–679. <https://doi.org/10.2214/AJR.04.1513>
- Diel J, Perlmutter S, Venkataraman N, Mueller R, Lane MJ, Katz DS (2000) Unenhanced helical CT using increased pitch for suspected renal colic: an effective technique for radiation dose reduction? *J Comput Assist Tomogr* 24(5):795–801. <https://doi.org/10.1097/00004728-200009000-00023>
- Heyer CM, Mohr PS, Lemburg SP, Peters SA, Nicolas V (2007) Image quality and radiation exposure at pulmonary CT angiography with 100- or 120-kVp protocol: prospective randomized study. *Radiology* 245(2):577–583. <https://doi.org/10.1148/radiol.2452061919>

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