Thieme

Value of vendor-agnostic deep learning image denoising in brain computed tomography: A multi-scanner study

Wertigkeit von Geräte-unabhängigem Deep-Learning Denoising in der Computertomographie: Eine Multiscanner-Studie

Authors

Christian Kapper[1], Lukas Müller[2], Andrea Kronfeld[1], Mario Alberto Abello Mercado[1], Sebastian Altmann[1], Nils Grauhan[1], Dirk Graafen[2], Marc A. Brockmann[1], Ahmed E. Othman[1]

Affiliations

1 Department of Neuroradiology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany
2 Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

Correspondence

Ahmed E. Othman
Department of Neuroradiology, University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, 55126 Mainz, Germany
ahmed.e.othman@googlemail.com

ABSTRACT

Purpose To evaluate the effect of a vendor-agnostic deep learning denoising (DLD) algorithm on diagnostic image quality of non-contrast cranial computed tomography (ncCT) across five CT scanners.

Materials and Methods This retrospective single-center study included ncCT data of 150 consecutive patients (30 for each of the five scanners) who had undergone routine imaging after minor head trauma. The images were reconstructed using filtered back projection (FBP) and a vendor-agnostic DLD method. Using a 4-point Likert scale, three readers performed a subjective evaluation assessing the following quality criteria: overall diagnostic image quality, image noise, gray matter-white matter differentiation (GM-WM), artifacts, sharpness, and diagnostic confidence. Objective analysis included evaluation of noise, contrast-to-noise ratio (CNR), signal-to-noise ratio (SNR), and an artifact index for the posterior fossa.

Results In subjective image quality assessment, DLD showed constantly superior results compared to FBP in all categories and for all scanners ($p < 0.05$) across all readers. The objective image quality analysis showed significant improvement in noise, SNR, and CNR as well as for the artifact index using DLD for all scanners ($p < 0.001$).

Conclusion The vendor-agnostic deep learning denoising algorithm provided significantly superior results in the subjective as well as in the objective analysis of ncCT images of patients with minor head trauma concerning all parameters compared to the FBP reconstruction. This effect has been observed in all five included scanners.

Key Points

- Significant improvement of image quality for 5 scanners due to the vendor-agnostic DLD
- Subjects were patients with routine imaging after minor head trauma
- Reduction of artifacts in the posterior fossa due to the DLD
- Access to improved image quality even for older scanners from different vendors

ZUSAMMENFASSUNG

Ziel Auswertung der Wirkung eines herstellerunabhängigen Deep Learning Denoising-Algorithmus (DLD) auf die diagnostische Bildqualität kontrastloser kranialer Computertomografie (ncCT) für fünf CT-Scanner im Vergleich.

Material und Methoden Diese retrospektive monozentrische Studie schloss ncCT-Daten von 150 konsekutiven Patienten (30 für jeden der fünf Scanner) ein, bei denen nach einem leichten Kopftrauma eine Routinebildgebung erfolgt war. Die Bilder wurden mittels gefilterter Rückprojektion (FBP) und ei-

ner herstellerunabhängigen DLD-Methode rekonstruiert. Anhand einer 4-Punkte-Likert-Skala führten drei Reader eine subjektive Bewertung durch, bei der die Qualitätskriterien allgemeine diagnostische Bildqualität, Bildrauschen, Differenzierung zwischen grauer und weißer Substanz (GM-WM), Artefakte, Bildschärfe und diagnostische Sicherheit bewertet wurden. Die objektive Analyse umfasste die Bewertung des Rauschens, des Kontrast-Rausch-Verhältnisses (CNR), des Signal-Rausch-Verhältnisses (SNR) und einen Artefaktindex für die Fossa cranii posterior.

Ergebnisse Bei der subjektiven Auswertung der Bildqualität zeigte DLD im Vergleich zu FBP in allen Bewertungskategorien und für alle Scanner konstant bessere Ergebnisse (p < 0,05) bei allen Readern. Die objektive Bildqualitätsanalyse zeigte bei allen Scannern eine signifikante Verbesserung des Rauschens, der SNR und der CNR sowie des Artefaktindexes durch das DLD (p < 0,001).

Schlussfolgerung Der herstellerunabhängige Deep Learning Denoising-Algorithmus lieferte im Vergleich zur FBP-Rekonstruktion bei allen Parametern sowohl in der subjektiven als auch in der objektiven Analyse deutlich bessere Ergebnisse für ncCT-Bilder von Patienten nach einem leichten Schädeltrauma. Dieser Effekt wurde bei allen fünf einbezogenen Scannern beobachtet.

Kernaussagen
- Hochsignifikante Verbesserung der Bildqualität für alle 5 Scanner durch das herstellerunabhängige DLD
- Eingeschlossen wurden Patienten mit Routinebildgebung nach leichtem Schädeltrauma
- Verringerung von Artefakten in der hinteren Schädelgrube durch das DLD
- Zugang zu verbesserter Bildqualität auch für ältere Geräte unterschiedlicher Hersteller möglich

Introduction

Due to fast accessibility and short examination time, non-contrast computed tomography (ncCT) has an outstanding importance, especially in the field of head trauma diagnostics because of its high sensitivity for the detection of intracranial hemorrhage [1, 2]. Yet, satisfying image quality, depiction of the gray and white matter, and thus reliable detection of small pathologies and lesions in cranial ncCT is rendered more difficult due to image noise and limited intrinsic differences in the brain parenchyma [3], especially while trying to leave the radiation dose at a reasonable level. With a constantly rising number of CT scans nationally and internationally [4, 5, 6], and thus increased accumulated radiation exposure and the risks of neoplasia that arise from it [7], new methods to reduce image noise, to increase image quality, and to enable a reduction in dose exposure for the patients have been introduced. Filtered back projection (FBP) has been a standard for CT image reconstruction for over 40 years due to its computational efficiency [8, 9]. However, it is limited in its potential for the improvement of image quality, dose, and noise reduction because of an increase in noise proportional to the inverse square root of the radiation dose [10]. A further step in attempts to improve CT imaging is iterative reconstruction (IR), which renders higher image quality and significant noise reduction compared to FBP [3, 11, 12] but has recently been challenged by new deep learning denoising (DLD) algorithms for CT image reconstruction, which are already showing superior results compared to FBP as well as IR [8]. Several studies have examined the effect of these new algorithms compared to conventional reconstruction methods, concentrating on specific body regions (e. g. head, lung, liver) using vendor-specific solutions [3, 13, 14, 15, 16, 17]. Yet, these denoising algorithms are limited to certain scanners from the same vendor and thereby exclude older scanners and scanners by others. Therefore, the evaluation of vendor-agnostic methods presents an opportunity to implement solutions to reduce noise and thus improve image quality for a large range of scanners independent of their age or producer. In first studies, the vendor-agnostic DLD we used for this study has already shown promising results for phantoms [18] as well as several body regions [19, 20, 21, 22, 23]. Yet, up to now, the important field of head CT has not been investigated using vendor-agnostic solutions. Therefore, the purpose of the study was to assess the ability of a vendor-agnostic DLD algorithm to denoise ncCT images of the brain, to improve the soft-tissue contrast compared to the corresponding FBP reconstructed series, and to produce comparable results for various CT scanners of different brands.

Materials and Methods

Subjects

In total, 150 patients (67 women, 83 men, median age: 79, age range: 19–95) were enrolled. All of them underwent an examination with the standard non-contrast head protocol (▶ **Table 1**) in our institutions after a minor head trauma. The examinations were executed on 5 different scanners (30 consecutive patients per scanner). Underage patients, patients with polytrauma and after surgical and interventional in-brain procedures were excluded beforehand.

Scanners and acquisition parameters

Five different scanners were included, Aquilion Precision (Canon Medical Systems Corporation, Japan; Scanner 1), Aquilion 32 (Canon Medical Systems Corporation, Japan; Scanner 2), iCT 256 (Philips Healthcare, Cleveland, OH, USA; Scanner 3), Brilliance 16 (Philips Healthcare, Cleveland, OH, USA; Scanner 4), and Brilliance 64 (Philips Healthcare, Cleveland, OH, USA; Scanner 5). All the images in all the scanners were generated using a tube voltage of 120 kV and a tube current depending on automatic modulation, which resulted in a tube current (± standard deviation [SD]) of 262.70 ± 28.43 mA for Scanner 1, 176.00 ± 12.00 mA for Scanner 2, 169.80 ± 8.19 mA for Scanner 3, 128.00 ± 8.18 mA for Scanner

▶ **Table 1** Scanning parameters for scanners 1–5.

	Scanner 1	Scanner 2	Scanner 3	Scanner 4	Scanner 5
Scanner	Canon/Toshiba Aquilion Precision	Canon/Toshiba Aquilion 32	Philips iCT 256	Philips Brilliance 16	Philips Brilliance 64
Manufacturer	Canon Medical Systems Corporation; Japan	Canon Medical Systems Corporation; Japan	Philips Healthcare, Cleveland, OH, USA	Philips Healthcare, Cleveland, OH, USA	Philips Healthcare, Cleveland, OH, USA
Scan mode	Axial	Axial	Axial	Axial	Axial
Tube voltage [kV]	120	120	120	120	120
Mean tube current [mA] ± SD	262.70 ± 28.43	176.00 ± 12.00	169.80 ± 8.19	128.00 ± 8.18	227.60 ± 7.72
Beam collimation	0.5 × 80 mm	0.5 × 32 mm	0.625 × 16 mm	4 × 4.5 mm	0.625 × 16 mm
Rotation time [s]	1.0	0.75	0.75	0.75	0.5
Field of view	210	210	Depending on patient	Depending on patient	Depending on patient
Matrix	512 × 512	512 × 512	512 × 512	512 × 512	512 × 512
Slice thickness	1 mm	1 mm	1 mm	1 mm	1 mm

4, and 227.60 ± 7.72 for Scanner 5. All scanners used a 512 × 512 matrix. For more details, see ▶ **Table 1**.

Denoising method

FBP was chosen as the standard reconstruction mode for this study to enable comparability. IR was not available on scanners 2, 4, and 5 and was therefore excluded. A commercially available and FDA approved DLD (ClariCT.AI, ClariPi, South Korea) [24] was applied to the FBP reconstructed images. It was developed as a denoising solution using a U-net based convolutional neural net [19], trained by taking a noise-added CT image as an input to produce an original CT image as an output. DLD was trained with diverse vendor-specific low-dose CT images from different vendors to acquire a generalized learning and vendor-agnostic denoising capability [10]. The training dataset consisted of over one million CT images encompassing 2,100 different combinations of scan and reconstruction conditions including varying kVp, mAs, automatic exposure control, slice thickness, contrast enhancement, and convolution kernels with 24 scanner models of 4 different CT manufacturers (GE Healthcare, Siemens Healthineers, Philips Healthcare, and Canon Medical): 80 % of the dataset was used for model training while the remaining 20 % were used for validation. The DLD focuses exclusively on tackling the task of noise reduction in the image domain, distinguishing it from the broader scope of CT deep learning reconstruction (DLR). The DLR was designed to handle the entire reconstruction process, which encompasses photon starvation compensation, beam hardening correction, the transformation of sinogram to CT images, and noise reduction [25]. In comparison, the DLD specifically targets noise reduction, honing its network capacity to this single task. This focused approach of the DLD ensures that it does not introduce the kinds of distortions or artifacts that are occasionally observed with DLR [26].
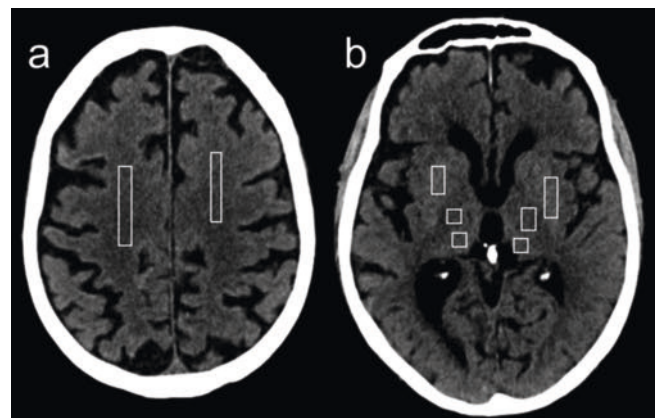
▶ **Fig. 1** Examples of the positioning of the ROIs on axial CT images of the **a** semioval center and **b** basal ganglia for objective analysis.

Image quality assessment
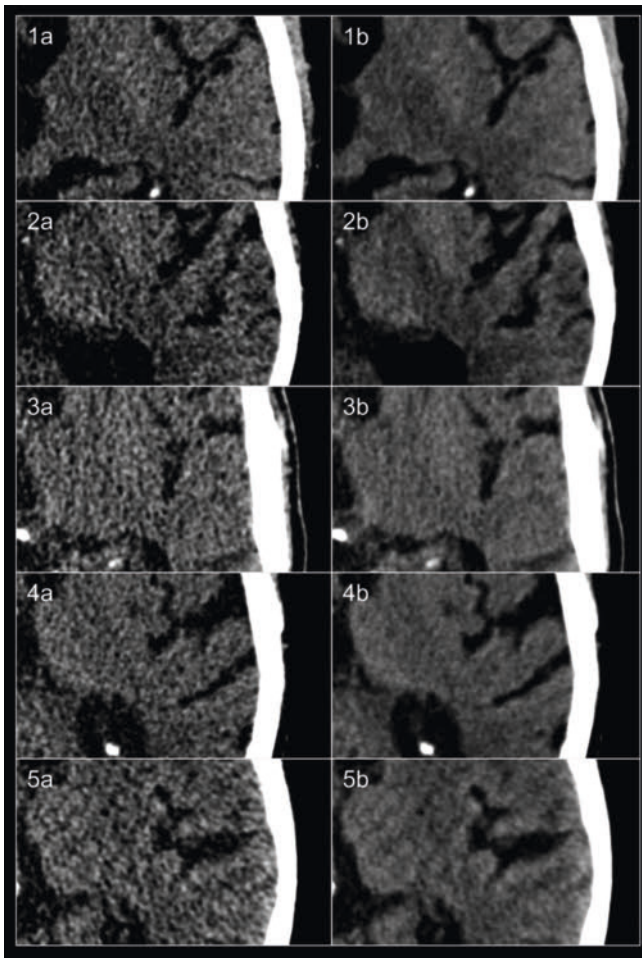
Subjective image evaluation

Subjective image quality was assessed independently by three radiologists (Reader 1: SA, Reader 2: MAM: both 6 years of experience in the field of neuroradiology, Reader 3: NG: 4 years), each trained with ten separate series selected beforehand as examples that were not included in the study. The readers were blinded to the reconstruction mode, scanner type, and the results of the objective image quality analysis. Parameters were overall image quality (general appearance), image noise, sharpness (depiction of the boundaries of the brain), brain structures (gray matter-white matter differentiation), overall artifacts and diagnostic confidence (ability to give a reliable diagnosis). They were evaluated using the following 4-point Likert scale: 1 = non-diagnostic; 2 = suboptimal but diagnostic; 3 = good; 4 = excellent.

▲ **Table 2** Medians and interquartile range for all 3 readers separated by FBP vs. DLD and by every scanner.

Reader 1

	Scanner 1 FBP	Scanner 1 DLD	Scanner 2 FBP	Scanner 2 DLD	Scanner 3 FBP	Scanner 3 DLD	Scanner 4 FBP	Scanner 4 DLD	Scanner 5 FBP	Scanner 5 DLD
Image quality	2 (2–3)	4 (3–4)	2 (2–2)	3 (3–4)	2 (2–2)	3 (3–4)	2 (2–2)	3 (3–3)	2 (2–2)	3 (3–4)
Noise	3 (2–3)	4 (3–4)	2 (2–2)	4 (3–4)	2 (2–2.75)	4 (3–4)	2 (2–3)	3 (3–4)	2 (2–2.75)	4 (3–4)
Sharpness	2 (2–3)	3 (3–4)	2 (2–2)	3 (3–4)	2 (2–2)	3 (3–3.75)	2 (2–2.75)	3 (3–4)	2 (2–3)	3 (3–4)
Brain structures	2 (2–3)	4 (3–4)	2 (2–2)	3.5 (3–4)	2 (2–2)	3 (2–4)	2 (2–3)	3 (3–4)	2 (1–2)	3 (3–4)
Artifacts	3 (3–4)	4 (3–4)	2 (2–3)	4 (4–4)	3 (2–3)	4 (4–4)	3 (2–3)	4 (3–4)	3 (2–3)	4 (4–4)
Diagnostic confidence	3 (2–3)	4 (3–4)	2 (2–2.75)	4 (3–4)	2 (2–3)	3 (3–4)	2 (2–3)	3 (3–4)	2 (2–3)	3.5 (3–4)

Reader 2

	Scanner 1 FBP	Scanner 1 DLD	Scanner 2 FBP	Scanner 2 DLD	Scanner 3 FBP	Scanner 3 DLD	Scanner 4 FBP	Scanner 4 DLD	Scanner 5 FBP	Scanner 5 DLD
Image quality	2 (2–3)	4 (3–4)	2 (2–2)	3 (3–4)	2 (1.25–2)	3 (3–4)	2 (2–2.75)	3 (3–4)	2 (2–2)	3 (3–4)
Noise	2 (2–2)	3 (3–3)	2 (2–2)	3 (3–3)	2 (1–2)	3 (3–3)	2 (2–2)	3 (3–3)	2 (1.25–2)	3 (3–3)
Sharpness	2 (2–3)	4 (3–4)	2 (2–2)	3.5 (3–4)	2 (2–2)	3 (3–4)	2 (2–2)	3 (3–4)	2 (2–2)	3.5 (3–4)
Brain structures	2 (2–2)	4 (3–4)	2 (2–2)	3 (3–4)	2 (2–2)	3 (3–4)	2 (2–2.75)	3 (3–4)	2 (2–2)	3 (3–4)
Artifacts	2.5 (2–3)	3.5 (3–4)	2 (2–3)	3.5 (3–4)	2 (2–2)	3 (3–3)	2 (2–3)	4 (3–4)	2 (2–2.75)	3.5 (3–4)
Diagnostic confidence	2 (2–3)	4 (3–4)	2 (2–2)	3.5 (3–4)	2 (2–2)	3 (3–4)	2 (2–3)	3 (3–4)	2 (2–2)	4 (3–4)

Reader 3

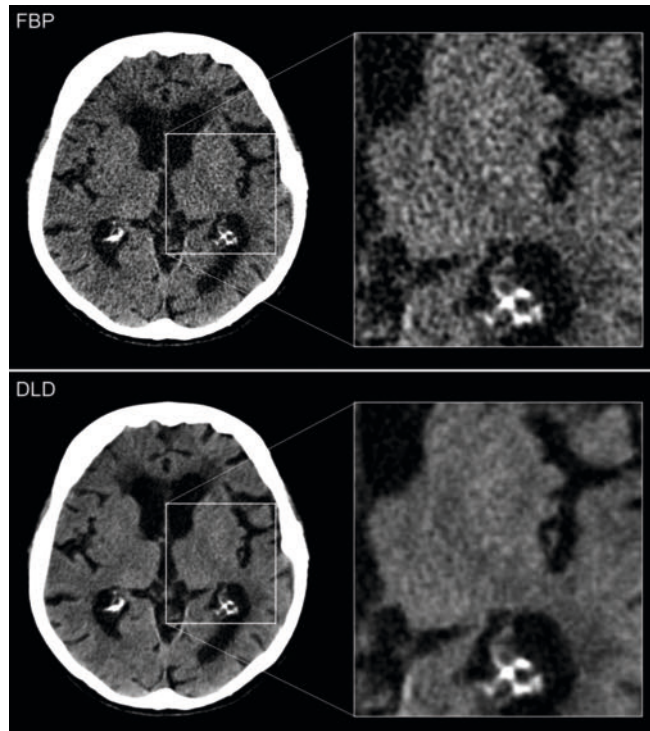|	Scanner 1 FBP	Scanner 1 DLD	Scanner 2 FBP	Scanner 2 DLD	Scanner 3 FBP	Scanner 3 DLD	Scanner 4 FBP	Scanner 4 DLD	Scanner 5 FBP	Scanner 5 DLD
Image quality	3 (2–3)	4 (4–4)	2 (2–2.75)	3 (3–4)	2 (1.25–2)	3 (3–3.75)	3 (2–3)	3 (3–4)	2 (1–2)	3.5 (3–4)
Noise	3 (2–3)	4 (3–4)	2 (2–2)	4 (3–4)	2 (1.25–2)	3.5 (3–4)	3 (2–3)	4 (3–4)	2 (1–2)	4 (3–4)
Sharpness	3 (3–3.75)	4 (4–4)	3 (3–3)	3 (2–4)	3 (2–3)	3 (3–3)	3 (2–3)	3 (3–4)	3 (2–3)	3 (3–4)
Brain structures	3 (2–3)	4 (4–4)	2 (2–3)	4 (3–4)	2 (2–2)	3 (3–3.75)	2 (2–3)	4 (3–4)	2 (2–2)	3.5 (3–4)
Artifacts	3 (3–4)	4 (4–4)	3 (3–4)	4 (4–4)	3 (3–3)	4 (4–4)	3 (3–4)	4 (4–4)	3 (3–3)	4 (4–4)
Diagnostic confidence	3 (2–3)	4 (4–4)	2 (2–3)	3 (3–4)	2 (2–2)	3 (3–3.75)	3 (2–3)	3 (3–4)	2 (1–2)	3 (3–4)
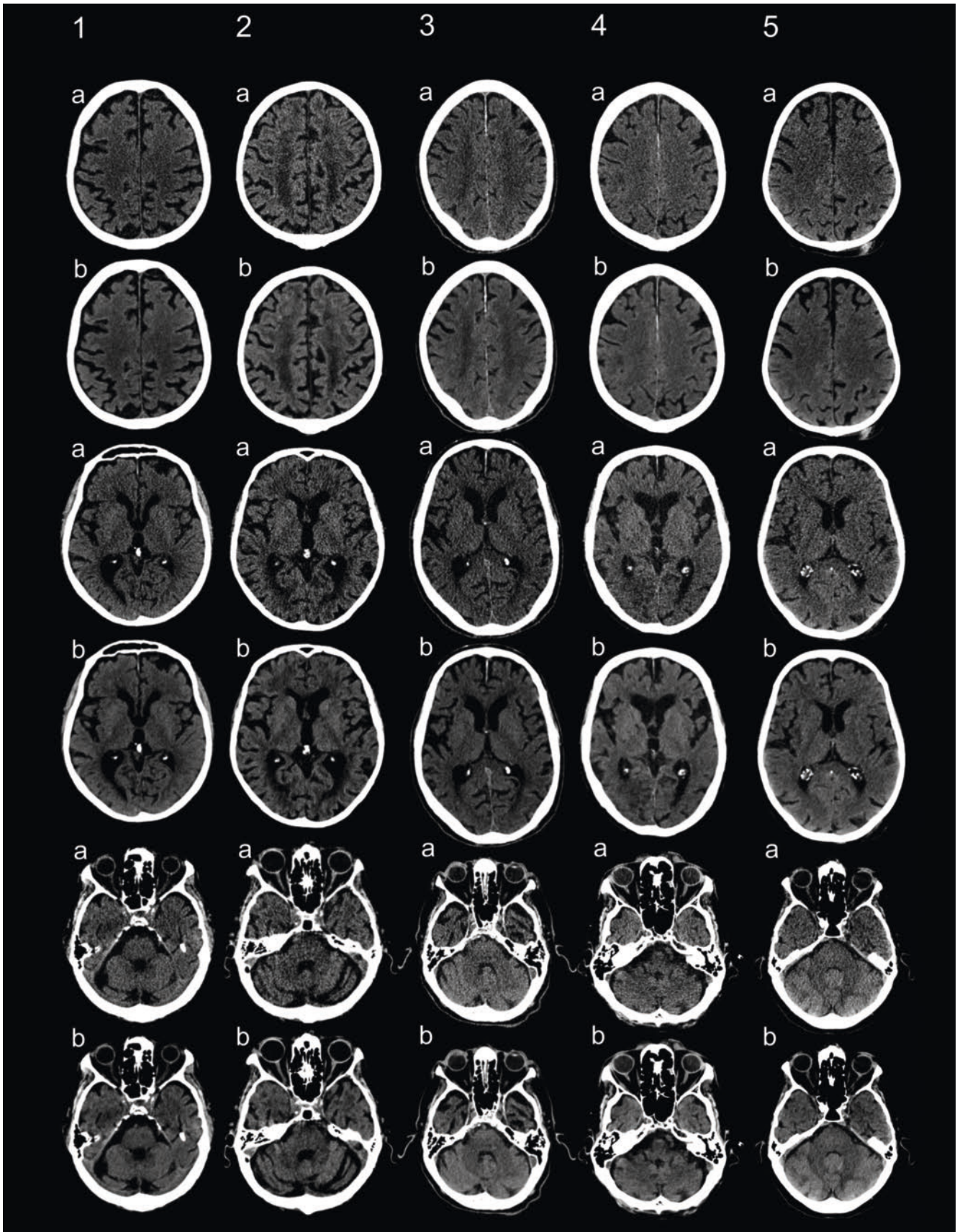
▶ **Fig. 2** Direct visual comparison of all 5 scanners (columns 1–5) and of **a** FBP vs. **b** DLD in 3 levels: semioval center, level of the basal ganglia and posterior fossa. 1 = Scanner 1, 2 = Scanner 2, 3 = Scanner 3, 4 = Scanner 4, 5 = Scanner 5.

> **Fig. 3** Zoomed in comparison of all 5 scanners (1 = Scanner 1, 2 = Scanner 2, 3 = Scanner 3, 4 = Scanner 4, 5 = Scanner 5) and of **a** FBP vs. **b** DLD.

> **Fig. 4** Patient example: 91-year-old female patient, control CT on scanner 5 after head trauma, anticoagulants in medication. Zoomed in comparison **a** FBP vs. **b** DLD.

Objective image evaluation

To evaluate noise, signal-to-noise ratio (SNR), and contrast-to-noise ratio (CNR) of white matter (WM) and gray matter (GM) regions, we used rectangular regions of interest (ROIs). The ROIs were drawn in the semioval center (for the measurement of SNR), the thalamus, the lentiform nucleus, and the adjacent white matter in the internal capsule (for the measurement of CNR), for each location, where possible, left and right (▶ **Fig. 1**). The exact position of the ROIs was copied automatically for the corresponding FBP and DLD images. The algorithm to determine the local noise is based on the work of Anam et al. [27] and Chun et al. [28] and is described by Altmann et al. in Ref. [29]. The goal is to determine noise that is not affected by tissue structures or gray value trends. Therefore, a sliding window of 5 mm x 5 mm is applied to the previously selected ROI. In this way, borders in the ROI were detected by a Sobel operator. If no anatomical border can be found at the present sliding window position, the region is detrended by a 2 D polynomial fit of second order [30] and the standard deviation (SD) of the gray values is calculated. The minimum standard deviation of the whole ROI out of all window positions was defined as noise. The noise and the mean CT density

[HU] in the corresponding square were used for the calculation of SNR by dividing CT density by SD. CNR was calculated with the following formula:

$$CNR = \frac{CT\ density_{GM} - CT\ density_{WM}}{noise_{WM}}$$

Furthermore, we evaluated a reduction of artifacts in the posterior fossa by setting ROIs of $200 \pm 3\ mm^2$ in the image with the most prominent artifacts. The SD in the ROI was defined as the artifact index [13].

Radiation dose

Radiation dose descriptors reported from each scanner were computed tomography dose index ($CTDI_{vol}$) and dose length product (DLP); the mean effective dose was estimated from the DLP using the conversion factor of 0.002 [mSv/mGycm] following Shrimpton et al. [31]. $CTDI_{vol}$: Scanner 1: 38.5 ± 4.1 mGy, Scanner 2: 45.3 ± 3.1 mGy, Scanner 3: 45.7 ± 6.7 mGy, Scanner 4: 46.2 ± 2.3 mGy, Scanner 5: $58,3 \pm 2.0$ mGy; DLP: Scanner 1: 760 ± 90 mGy*cm, Scanner 2: 783 ± 70 mGy*cm, Scanner 3: 894 ± 160 mGy*cm, Scanner 4: $830 +/ 100$ mGy*cm, Scanner 5: 974 ± 56 mGy*cm; mean effective dose ± SD: Scanner 1: 1.5 ± 0.2 mSv, Scanner 2: 1.6 ± 0.15 mSv, Scanner 3: 1.8 ± 0.3 mSv, Scanner 4: 1.7 ± 0.2 mSv, Scanner 5: 2 ± 0.1 mSv.

▲ **Table 3** Results of the objective image quality evaluation. LN = lentiform nucleus; Th = thalamus.

| Reconstruction		Scanner 1		Scanner 2		Scanner 3		Scanner 4		Scanner 5	
		FBP	DLD	FBP	DLD	FBP	DLD	FBP	DLD	FBP	DLD
Noise [mean and SD]		3.9 / 0.4	1.8 / 0.2	3.2 / 0.3	1.5 / 0.2	4.3 / 0.8	2.0 / 0.4	3.5 / 0.4	1.6 / 0.2	4.0 / 0.6	1.9 / 0.3
p-values		<0.001		<0.001		<0.001		<0.001		<0.001	
SNR [mean and SD]		271.7 / 25.0	575.3 / 55.5	328.8 / 35.6	717.0 / 71.4	252.6 / 45.7	549.1 / 100.6	307.8 / 35.4	652.3 / 74.4	266.6 / 41.8	572.1 / 84.3
p-values		<0.001		<0.001		<0.001		<0.001		<0.001	
CNR [mean and SD, median]	LN right	2.1 / 0.7	4.6 / 1.1	3.0 / 1.0	6.3 / 1.8	2.2 / 1.0	4.5 / 1.7	2.3 / 0.8	4.6 / 1.5	1.8 / 0.7	3.6 / 1.3
	median	2.2	4.6	2.9	6.2	2.0	4.2	2.4	4.4	1.8	3.2
	LN left	2.3 / 0.6	4.6 / 1.1	3.0 / 0.9	6.2 / 1.6	2.3 / 0.9	5.1 / 2.4	2.4 / 0.9	4.9 / 1.8	2.1 / 0.9	4.1 / 1.0
	median	2.2	4.7	2.9	6.2	2.1	4.4	2.1	4.4	2.1	4.0
	Th right	1.8 / 0.6	3.8 / 1.1	2.6 / 0.8	5.6 / 1.3	2.1 / 1.0	4.5 / 2.0	2.5 / 0.8	4.9 / 1.6	1.9 / 0.7	3.9 / 1.2
	median	1.8	3.9	2.5	5.4	1.9	4.0	2.4	4.4	2.0	3.9
	Th left	1.9 / 0.6	3.8 / 1.1	2.7 / 0.8	5.3 / 1.5	2.0 / 0.9	4.7 / 2.2	2.4 / 0.7	4.9 / 1.6	1.9 / 0.6	3.6 / 1.2
	median	1.8	3.6	2.5	5.2	2.0	4.1	2.3	4.7	1.9	3.4
Artifact index [mean and SD]		9.1 / 1.3	5.8 / 1.2	9.4 / 0.8	5.6 / 0.6	11.2 / 1.9	6.7 / 1.6	9.6 / 1.2	6.1 / 1.2	10.3 / 1.6	6.0 / 1.1
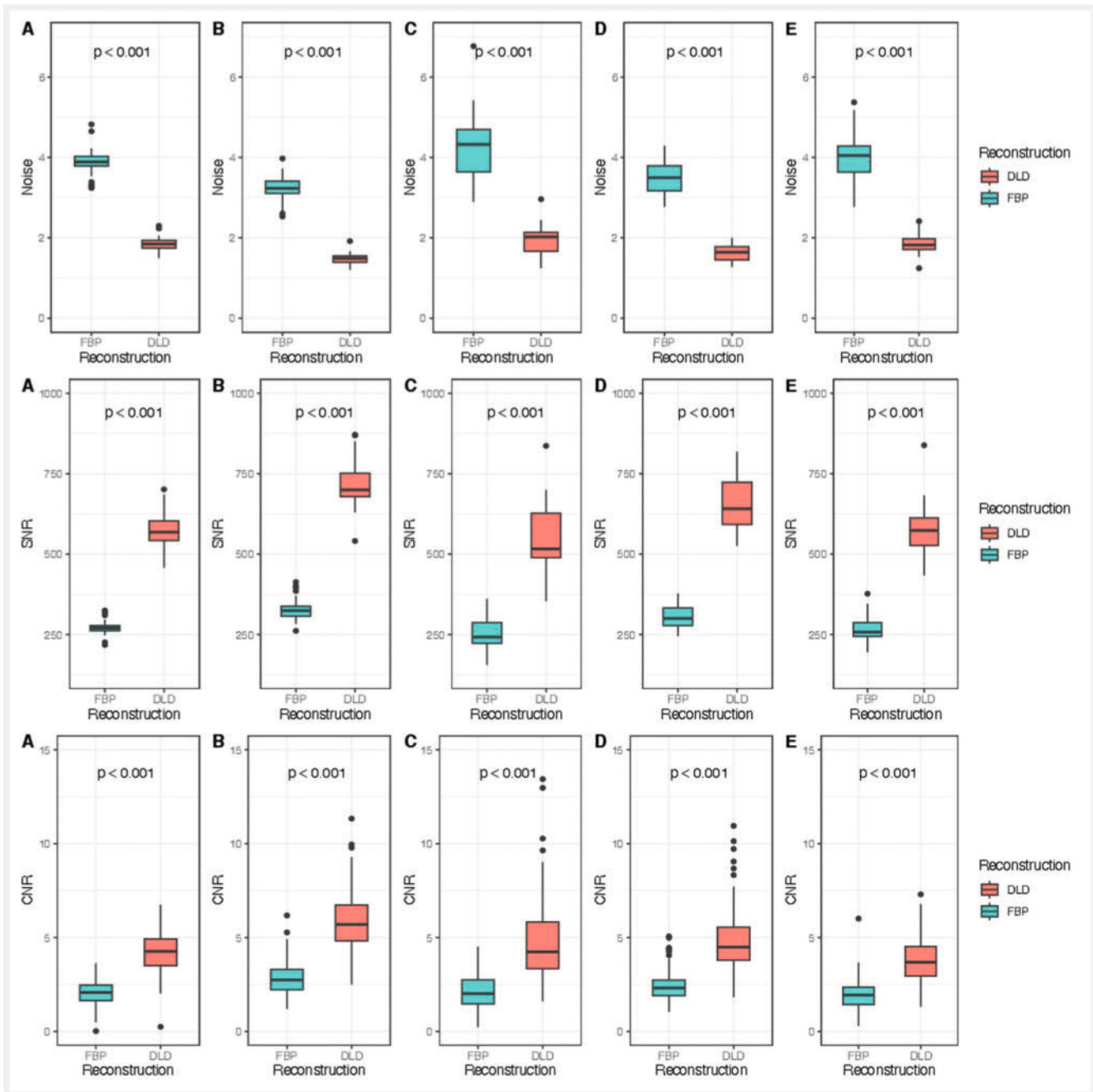
▶ **Fig. 5** Boxplot and significance levels for noise, SNR, and CNR compared between all 5 scanners: **A** = Scanner 1, **B** = Scanner 2, **C** = Scanner 3, **D** = Scanner 4, **E** = Scanner 5, FBP vs. DLD; for CNR, mean values across the lentiform nucleus, thalamus, and left and right were used for clarity.

Statistical analysis

All statistical analyses were performed with R 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria.).

For the data raised in the subjective quality analysis, we performed a Wilcoxon-signed rank test where a p-value of < 0.05 was considered significant. Furthermore, the inter-observer variability was determined with intraclass correlation. An intraclass correlation coefficient (ICC) of up to 0.2 was considered to be very poor agreement, 0.2–0.4 poor agreement, 0.4–0.6 fair agreement, 0.6–0.8 good agreement, and > 0.8 almost perfect agreement.

The data acquired in the objective analysis showed no normal distribution applying the Shapiro-Wilk test. Therefore, comparative analyses were performed using the Wilcoxon signed-rank test. A p-value of < 0.05 was considered statistically significant. For the comparison of the effect size of the DLD on the images from the different manufacturers, we used Cohen's d regarding noise, SNR, and CNR. A d-value of 0.2–0.5 was deemed a small effect, between 0.5 and 0.8 a medium one, and > 0.8 was deemed a high effect.
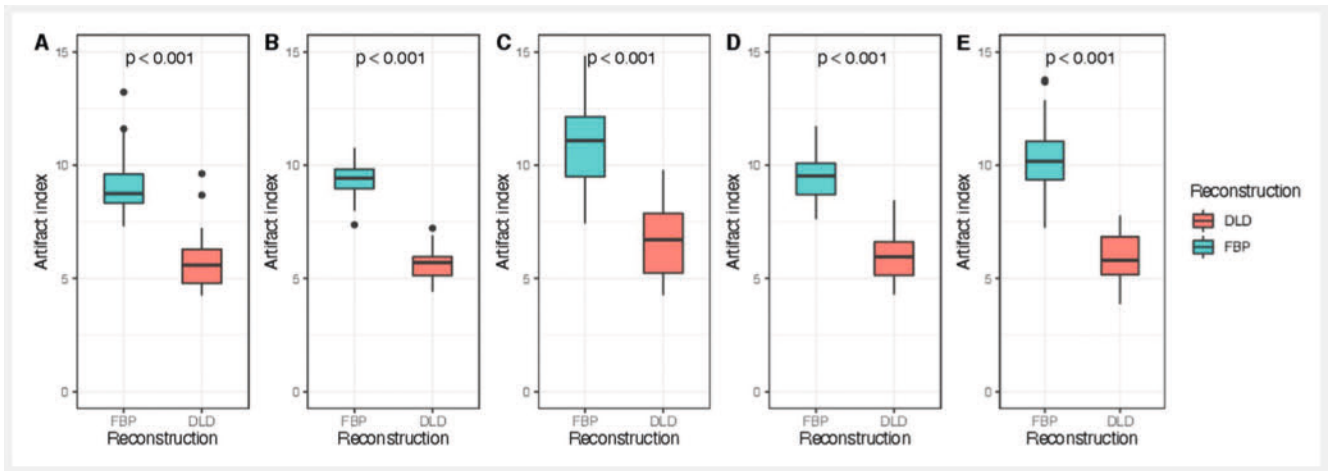
► **Fig. 6** Boxplot and significance levels for the artifact index compared between all 5 scanners: **A** = Scanner 1, **B** = Scanner 2, **C** = Scanner 3, **D** = Scanner 4, **E** = Scanner 5, FBP vs. DLD.

Results

Subjective quality analysis

The subjective quality analysis showed significantly superior values for the DLD compared to the FBP reconstructions across all scanners and all readers for overall image quality, noise, sharpness, brain structures, artifacts, and diagnostic confidence (each $p < 0.001$). DLD images were rated good to excellent, whereas values for FBP ranged from non-diagnostic to excellent with the large majority being suboptimal but diagnostic. Scanner 1 tended to have the best ratings of the five scanners included across all assessed criteria. ► **Table 2** gives a detailed overview of the reading results subdivided into the three readers and five scanners. Inter-reader agreement was good for image quality (ICC = 0.661), noise (ICC = 0.624), brain structures (ICC = 0.66), and diagnostic confidence (ICC = 0.616), and fair for sharpness (ICC = 0.407) and for artifacts (ICC = 0.446). Patient examples are given in ► **Fig. 2**, ► **Fig. 3** and ► **Fig. 4**. Due to the small number of pathologies included as described above, the effect of the DLD on the depiction of these pathologies was not evaluated.

Objective quality analysis

Compared to the FBP images, noise was reduced and SNR as well as CNR were improved significantly ($p < 0.001$) when the DLD was applied. This has been observed consistently for every scanner independent of the different positions of the ROIs (internal capsule for SNR, lentiform nucleus, thalamus for CNR) (► **Table 3**, ► **Fig. 5**). The artifact index evaluated in the posterior fossa showed a significant reduction in all scanners (each $p < 0.001$) for the DLD compared to FBP (► **Table 3**, ► **Fig. 6**). Cohen's d showed a large effect size for all the scanners with respect to noise, SNR, and CNR (Scanner 1 > 4.2; Scanner 2 > 4.3; Scanner 3 > 1.5; Scanner 4 > 3.2; Scanner 5 > 1.6 for noise, SNR and CNR) (► **Table 4**). Yet, scanners 1, 2 and 4 seemed to profit most from the DLD.

► **Table 4** Effect sizes of the DLD for the 5 scanners (Cohen's d). A value of d > 0.8 was deemed a high effect.

	Noise	**SNR**	**CNR**
Scanner 1	4.2	4.4	4.2
Scanner 2	4.9	5.1	4.3
Scanner 3	1.5	1.6	1.6
Scanner 4	3.2	3.3	3.3
Scanner 5	1.6	2.1	2.1

Discussion

The purpose of this study was to evaluate the effect of a vendor-agnostic DLD solution on image quality compared to FBP reconstruction across different CT scanners by different manufacturers in the field of brain ncCT imaging.

Regarding the results of the subjective as well as the objective image quality assessment, the addition of DLD to FBP significantly outperformed FBP only, providing a reduction of noise and of the artifact index in the posterior fossa and higher image quality on every scanner.

We find our results to be in line with existing studies concerning the use of DLD algorithms for brain ncCT [3, 13, 14, 15, 32] (vendor-specific or not commercially available [32]) as well as for examinations of other regions of the body [10, 16, 17, 19, 20, 21, 22, 23] (vendor-specific and vendor-agnostic). All of them managed to detect superior image quality and noise reduction throughout. Kim et al. [13] reported a reduction in artifacts – which we also found – in the posterior fossa, all using DLD compared to FBP and/or compared to IR. We did not include IR in this study, because it was not available for every scanner, especially not for the older ones. Also, it was important for us to guarantee the comparability between the five scanners (for which FBP reconstructed images were a common base) in order to be able to de-

termine whether a large variety of scanners can profit from the use of vendor-agnostic DLD solutions. The results obtained in the subjective and objective analysis seem to support this thesis. Yet, in this study, we only included preexisting CT images of the brain that were acquired using the standard dose protocols for head trauma. Looking at the effect size and the quality gained thanks to the DLD, that leaves the possibility for further investigation in dose reduction and application of vendor-agnostic DLD methods on low and ultra-low dose cranial CT images, as was already shown for other regions of the body, for example, the lumbar spine [20]. Due to superior image quality showing significantly superior gray matter-white matter differentiation, pathologies or those with slight differences in density in the brain parenchyma might be detected more easily and reliably even in non-contrast brain CT examinations. For example, neoplasia and strokes [32] come to mind. These changes in the depiction of parenchymal contrast could also influence subjective scoring systems like ASPECTS (Alberta stroke program early CT score), which depend essentially on the visibility of those slight differences between gray and white matter. To determine to which degree DLDs affect their precision, further investigation is needed. Based on the answer to our main question in this study, an implication for practical use can be envisioned: The vendor-agnostic DLD methods allow for the possibility to have just one denoising engine in hospitals or radiological centers with more than one scanner, scanners from different vendors, and older scanners, while providing reliably good image quality.

This study has several limitations. The groups compared in this retrospective study were relatively small with 30 patients per scanner, and we only compared scanners by two of the leading companies in CT imaging. Due to the lack of IR in some of the scanners, we were not able to include this reconstruction method. The study focused on image quality assessment in a cohort of patients with a very low likelihood of pathological findings. Pathologies were therefore not analyzed separately. Apart from that, due to the retrospective study design, we face a potential selection bias and, by only using preexisting CT examinations with a fixed tube voltage of 120 kV, we are only able to speculate about the effectiveness of this DLD for dose reduction across the five scanners even though a benefit for low-dose CT images of the spine has already been proven [20]. These aspects need to be evaluated in further investigations. Sharpness as well as artifacts showed only fair agreement, which could be attributed to individual internal standards of the readers for the respective parameters.

The use of DLD in a clinical setting requires certain computational infrastructure (especially GPU).

Conclusion

The vendor-agnostic deep learning denoising algorithm provided significantly superior results in the subjective as well as in the objective analysis of ncCT images of patients with minor head trauma concerning all parameters compared to the FBP reconstruction. This effect has been observed in all five included scanners.

Clinical relevance

- Due to the DLD, a significant improvement in image quality was achieved compared to FBP.
- Vendor-agnostic DLD methods make it possible to have just one denoising engine in hospitals or radiological centers with more than one scanner, scanners from different vendors, and older scanners, while providing reliably good image quality.
- Further investigation in the field of dose reduction and regarding specific pathologies is conceivable.

Conflict of Interest

Ahmed E. Othman has been a consultant for ClariPi for the past three years. Christian Kapper received technical support in setting up ClariCT. AI from ClariPi employees.

References

[1] Neifert SN, Chapman EK, Martini ML et al. Aneurysmal Subarachnoid Hemorrhage: the Last Decade. Transl Stroke Res 2021; 12 (3): 428–446. doi:10.1007/s12975-020-00867-0

[2] Currie S, Saleem N, Straiton JA et al. Imaging assessment of traumatic brain injury. Postgrad Med J 2016; 92: 41–50. doi:10.1136/postgradmedj-2014-133211

[3] Alagic Z, Diaz Cardenas J, Halldorsson K et al. Deep learning versus iterative image reconstruction algorithm for head CT in trauma. Emerg Radiol 2022; 29 (2): 339–352

[4] Schöckel L, Jost G, Seidensticker P et al. Developments in X-Ray Contrast Media and the Potential Impact on Computed Tomography. Invest Radiol 2020; 55 (9): 592–597. doi:10.1097/RLI.0000000000000696

[5] Dieckmeyer M, Sollmann N, Kupfer K et al. Computed Tomography of the Head. Clin Neuroradiol 2023; 33: 591–610. doi:10.1007/s00062-023-01271-5

[6] Bos D, Guberina N, Zensen S et al. Radiation Exposure in Computed Tomography. Dtsch Arztebl Int 2023; 120 (9): 135–141. doi:10.3238/arztebl.m2022.0395

[7] Miglioretti DL, Johnson E, Williams A et al. The Use of Computed Tomography in Pediatrics and the Associated Radiation Exposure and Estimated Cancer Risk. JAMA Pediatr 2013; 167 (8): 700–707. doi:10.1001/jamapediatrics.2013.311

[8] Arndt C, Güttler F, Heinrich A et al. Deep Learning CT Image Reconstruction in Clinical Practice. Fortschr Röntgenstr 2021; 193 (3): 252–261. doi:10.1055/a-1248-2556

[9] Kataria B, Nilsson Althén J, Smedby Ö et al. Image Quality and Potential Dose Reduction using Advanced Modeled Iterative Reconstruction (ADMIRE) in Abdominal CT – A Review. Radiat Prot Dosimetry 2021; 195: 177–187

[10] Ahn C, Heo C, Kim JH. Combined low-dose simulation and deep learning for CT denoising: application in ultra-low-dose chest CT. Proc SPIE Int Soc Opt Eng 2019. doi:10.1117/12.2521539.short

[11] Cho YJ, Schoepf UJ, Silverman JR et al. Iterative Image Reconstruction Techniques: Cardiothoracic Computed Tomography Applications. J Thorac Imaging 2014; 29 (4): 198–208. doi:10.1097/RTI.0000000000000041

[12] Bodelle B, Wichmann JL, Scholtz JE et al. Iterative Reconstruction Leads to Increased Subjective and Objective Image Quality in Cranial CT in Patients With Stroke. AJR Am J Roentgenol 2015; 205 (3): 618–622

[13] Kim I, Kang H, Yoon HJ et al. Deep learning–based image reconstruction for brain CT: improved image quality compared with adaptive statistical iterative reconstruction-Veo (ASIR-V). Neuroradiology 2021; 63 (6): 905–912

[14] Sun J, Li H, Wang B et al. Application of a deep learning image reconstruction (DLIR) algorithm in head CT imaging for children to improve image quality and lesion detection. BMC Med Imaging 2021; 21 (1): 108. doi:10.1186/s12880-021-00637-w

[15] Oostveen LJ, Meijer FJA, de Lange F et al. Deep learning–based reconstruction may improve non-contrast cerebral CT imaging compared to other current reconstruction algorithms. Eur Radiol 2021; 31 (8): 5498–5506. doi:10.1007/s00330-020-07668-x

[16] Park C, Choo KS, Jung Y et al. CT iterative vs deep learning reconstruction: comparison of noise and sharpness. Eur Radiol 2021; 31 (5): 3156–3164. doi:10.1007/s00330-020-07358-8

[17] Jiang B, Li N, Shi X et al. Deep Learning Reconstruction Shows Better Lung Nodule Detection for Ultra-Low-Dose Chest CT. Radiology 2022; 303 (1): 202–212

[18] Choi H, Chang W, Kim JH et al. Dose reduction potential of vendor-agnostic deep learning model in comparison with deep learning-based image reconstruction algorithm on CT: a phantom study. Eur Radiol 2022; 32 (2): 1247–1255

[19] Hong JH, Park EA, Lee W et al. Incremental Image Noise Reduction in Coronary CT Angiography Using a Deep Learning-Based Technique with Iterative Reconstruction. Korean J Radiol 2020; 21 (10): 1165–1177

[20] Yeoh H, Hong SH, Ahn C et al. Deep Learning Algorithm for Simultaneous Noise Reduction and Edge Sharpening in Low-Dose CT Images: A Pilot Study Using Lumbar Spine CT. Korean J Radiol 2021; 22 (11): 1850–1857. doi:10.3348/kjr.2021.0140

[21] Park S, Yoon JH, Joo I et al. Image quality in liver CT: low-dose deep learning vs standard-dose model-based iterative reconstructions. Eur Radiol 2022; 32 (5): 2865–2874. doi:10.1007/s00330-021-08380-0

[22] Lee S, Choi YH, Cho YJ et al. Noise reduction approach in pediatric abdominal CT combining deep learning and dual-energy technique. Eur Radiol 2021; 31 (4): 2218–2226. doi:10.1007/s00330-020-07349-9

[23] Kolb M, Storz C, Kim JH et al. Effect of a novel denoising technique on image quality and diagnostic accuracy in low-dose CT in patients with suspected appendicitis. Eur J Radiol 2019; 116: 198–204

[24] FDA. Premarket Notification (device name ClariCT.AI): FDA. 2019. https://www.accessdata.fda.gov/cdrh_docs/pdf18/K183460.pdf

[25] Hsieh J, Liu E, Nett B et al. A new era of image reconstruction: TrueFidelity™, Technical white paper on deep learning image reconstruction. 2019. https://www.gehealthcare.de/-/jssmedia/files/truefidelity/truefidelity-white-paper-jb68676xx-doc2287426.pdf?rev = -1

[26] Nam JG, Ahn C, Choi H et al. Image quality of ultralow-dose chest CT using deep learning techniques: potential superiority of vendor-agnostic post-processing over vendor-specific techniques. Eur Radiol 2021; 31 (7): 5139–5147. doi:10.1007/s00330-021-07733-z

[27] Anam C, Arif I, Haryanto F et al. An Improved Method of Automated Noise Measurement System in CT Images. J Biomed Phys Eng 2021; 11 (2): 163–174. doi:10.31661/jbpe.v0i0.1198

[28] Chun M, Choi YH, Kim JH. Automated measurement of CT noise in patient images with a novel structure coherence feature. Phys Med Biol 2015; 60 (23): 9107–9122. doi:10.1088/0031-9155/60/23/9107

[29] Altmann S, Abello Mercado MA, Ucar FA et al. Ultra-High-Resolution CT of the Head and Neck with Deep Learning Reconstruction; Assessment of Image Quality and Radiation Exposure and Intraindividual Comparison with Normal-Resolution CT. Diagnostics (Basel) 2023; 13 (9): 1534

[30] Dobbins JT 3 rd, Samei E, Ranger NT et al. Intercomparison of methods for image quality characterization. II. Noise power spectrum. Med Phys 2006; 33 (5): 1466–1475. doi:10.1118/1.2188819

[31] Shrimpton PC, Jansen JT, Harrison JD. Updated estimates of typical effective doses for common CT examinations in the UK following the 2011 national review. Br J Radiol 2016; 89 (1057): 20150346. doi:10.1259/bjr.20150346

[32] Wong KK, Cummock JS, He Y et al. Retrospective study of deep learning to reduce noise in non-contrast head CT images. Comput Med Imaging Graph 2021; 94: 101996