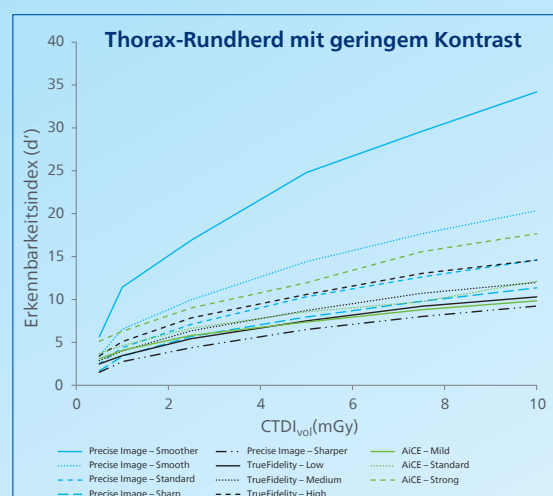


Precise Image: Ein aufgabenbasierter Vergleich der Bildqualität von mehreren Anbietern

Niels van der Werf, PhD – CT Clinical Scientist Europe – Philips Healthcare

Überblick

In diesem White Paper werden die Bildqualität der Deep-Learning-Rekonstruktion (DLR) mit Philips Precise Image objektiv quantifiziert und die Ergebnisse dieser KI-Rekonstruktion mit der DLR-Bildqualität anderer CT-Hersteller verglichen. Precise Image verbesserte die Bildqualität im Vergleich zur gefilterten Rückprojektion und iterativen Rekonstruktion signifikant. Darüber hinaus übertraf Precise Image die KI-Rekonstruktion der Mitbewerber in Bezug auf die Erkennbarkeitsindizes für Thorax-Rundherde mit geringem Kontrast.



Einführung

Seit Kurzem ist eine Deep-Learning-Rekonstruktionstechnik (DLR) namens Precise Image erhältlich, die die Grenzen klassischer CT-Rekonstruktionsmethoden überwindet. Während die klassische gefilterte Rückprojektion häufig stark verrauschte Bilder liefert und iterative Rekonstruktionstechniken (IR) das Rauschen zwar reduzieren, jedoch auf Kosten der Bildglätte bei unbekannten Texturen, bietet Precise Image die ideale Kombination.¹ Es liefert rauscharme Bilder, die Bildern aus der gefilterten Rückprojektion ähneln, jedoch ohne die Nachteile der gefilterten Rückprojektion oder iterativen Rekonstruktion.

Wenn es um die Bewertung der Bildqualität für verschiedene Rekonstruktionstechniken geht, sind herkömmliche Messgrößen wie Bildrauschen, Signal-Rausch-Verhältnis (SRV) und Kontrast-Rausch-Verhältnis (KVR) unzureichend.² Daher wurden andere, fortschrittlichere objektive Messgrößen für die Bildqualität vorgeschlagen, um die potenzielle Bildqualität in der klinischen Praxis besser zu beschreiben. Eine dieser Messgrößen ist der Erkennbarkeitsindex (d'), der die Fähigkeit von Radiologen bewertet, Läsionen unter Berücksichtigung von Faktoren wie Rauschen,

Auflösung und läsionsspezifischen Merkmalen genau zu erkennen. Dazu werden das Rauschleistungsspektrum (Noise-Power-Spectrum, NPS) und eine aufgabenbasierte Übertragungsfunktion (Task-based Transfer Function, TTF) herangezogen. Mit der NPS-Analyse lässt sich nicht nur das Rauschen, sondern auch die Textur und Granularität quantifizieren, während die TTF die Bildauflösung für bestimmte klinische Aufgaben bewertet. Zusammen sorgen diese Messgrößen dafür, dass Precise Image eine ausgezeichnete klinische Bildgebungsqualität liefert, die Diagnosesicherheit erhöht und die Behandlungsergebnisse verbessert.

Phantom-Setup

Zur objektiven Bewertung der Bildqualität von Precise Image wurde entsprechend den Methoden aus zwei kürzlich erschienenen Publikationen von Greffier et al. ein ACR 464 (CT ACR 464 Phantom, Gammex Technology) verwendet (**Abbildung 1**).^{3,4} Mit diesem Phantom wurden erweiterte quantitative Bildqualitätseigenschaften in Bezug auf Rauschen und Auflösung bewertet. Die Rauscheigenschaften wurden anhand des Rauschleistungsspektrums in 20 aufeinanderfolgenden Schichten eines homogenen Bereichs des Phantoms gemessen. Die Auflösungseigenschaften wurden mit einer TTF in zehn aufeinanderfolgenden Schichten für verschiedene aufgabenspezifische Kontraste, d.h. Luft (–1000 Hounsfield-Einheiten [HU]) und Polyethylen (–95 HU), geschätzt.

Auf einem Philips Incisive CT-System mit Softwareversion 5.01 (Philips, Best) wurden unter Verwendung eines Standard-Thorax-CT-Helikalkoll (Einblendung 64 x 0,625 mm; Rotationszeit 0,35 s; Pitch 1) mit einem Röhrenpotential von 120 kVp Rohdaten erfasst.⁴

Es wurden sechs verschiedene Dosen verwendet: 10, 7,5, 5, 2,5, 1 und 0,5 mGy. Dabei wurden feste Röhrenstromwerte (mA) verwendet. Die Rohdaten wurden auf einer Matrix von 512 x 512 Elementen mit einem 250-mm-Sichtfeld unter Verwendung aller drei verfügbaren Rekonstruktionsmethoden rekonstruiert: gefilterte Rückprojektion, iterative Rekonstruktion (iDose⁴, Philips) und DLR (Precise Image, Philips). Sowohl für die gefilterte Rückprojektion als auch für die iterative Rekonstruktion wurden der B- und YA-Filter eingesetzt. Für die iterative Rekonstruktion wurden die beiden Stufen 4 und 7 verwendet. Für die Philips DLR mit Precise Image wurden die Rohdaten mit allen verfügbaren Stufen (Smoother, Smooth, Standard, Sharp, Sharper) rekonstruiert, und zwar sowohl mit der Einstellung für Weichgewebe als auch für Lunge. In Übereinstimmung mit Greffier et al. wurde für alle Rekonstruktionen eine Schichtdicke von 1 mm mit einem überlappenden Schichtinkrement von 0,5 mm verwendet.⁴

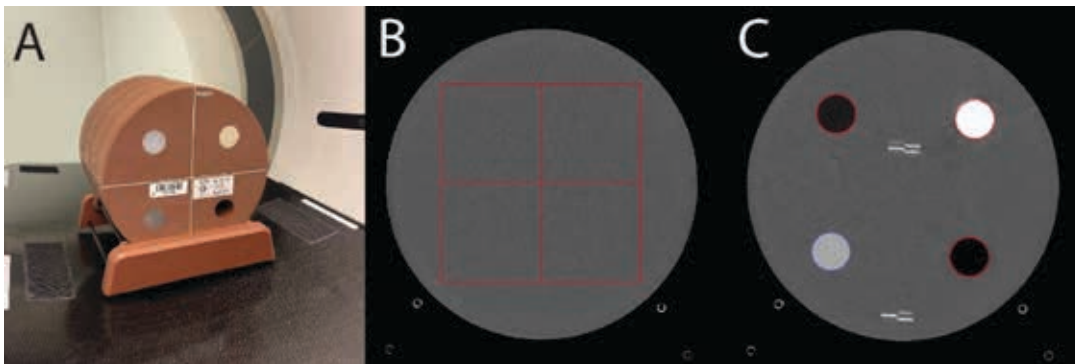


Abbildung 1: Überblick über das Phantom-Setup (A), die Schicht mit Rauschleistungsspektrum (B) und die Schicht mit aufgabenbasierter Übertragungsfunktion (C)

Aufgabenbasierte Bewertung der Bildqualität

Zur objektiven Bewertung der Bildqualität mit der iQMetrix-CT Software wurden drei Aufgaben definiert, die mit den von Greffier et al. definierten Aufgaben identisch sind.^{3,4} Dies waren folgende Aufgaben: Erkennung von Thoraxläsionen mit einem Durchmesser von 5 mm für einen Thorax-Rundherd mit geringem Kontrast im Weichgewebe, eine milchglasartige Dichteanhebung und eine pulmonale Läsion mit hohem Kontrast mit einem Kontrast von 50, 200 bzw. 950 HU. Für jede Erkennungsaufgabe wurde der Erkennbarkeitsindex anhand eines Beobachtermodells ohne Pre-Whitening mit Augenfilter berechnet.⁵⁻⁷ Für die beiden Aufgaben mit dem geringsten Kontrast wurden die TTF-Ergebnisse für Polyethylen verwendet, für die Aufgabe mit hohem Kontrast hingegen das TTF-Ergebnis für den Lufteinsatz. Alle Aufgaben wurden auf einer Matrix von 512 x 512 Elementen mit einer Pixelgröße von 0,488 x 0,488 mm definiert.² Für die Auswertung wurden ein Zoomfaktor von 1,5, ein Betrachtungsabstand von 500 mm und ein Sichtfeld von 500 mm verwendet. Für jede Aufgabe wurden die Ergebnisse von Precise Image mit den Ergebnissen von Greffier et al. für die DLR von Canon (AiCE) und GE (TrueFidelity™) aus dieser Publikation verglichen.³

Für eine vollständige Auswertung der Ergebnisse ist es wichtig, die Beziehung zwischen dem Erkennbarkeitsindex (d') und der ihm zugrunde liegenden ROC-Kurve (Receiver Operating Characteristic) zu verstehen.⁸ Die ROC-Kurve bildet die Wahrscheinlichkeit richtig positiver Erkennungen des untersuchten Objekts bzw. der untersuchten Läsion im Vergleich zu falsch positiven Erkennungen (Fälle, in denen der Beobachter entscheidet, dass das Objekt vorhanden ist, obwohl es in Wahrheit nicht vorhanden ist) ab. Dabei wird eine Entscheidungsschwelle über einen Bereich möglicher Werte variiert. Mit einem d' -Wert wird diese 2D-ROC-Kurve in einen 1D-Einzelwert umgerechnet, wobei Gaußsche Verteilungen der Entscheidungsvariablen angenommen werden. Ein d' -Wert = 3,5 entspricht einer AUC (Fläche unter der ROC-Kurve) = 0,9932, was bedeutet, dass der Beobachter das Objekt in 99,32% der Fälle richtig erkennt. Der schlechteste AUC-Wert von 0,5 bedeutet, dass der Beobachter nicht besser abschneidet als bei willkürlichem Raten, was $d' = 0$ entspricht. Somit stellen alle d' -Werte über etwa 3,5 keine signifikant bessere Erkennbarkeit mehr dar (da die Objekterkennung bei $d' \approx 3,5$ bereits nahezu perfekt ist). d' -Werte über 3,5 spiegeln meist eine Rauschunterdrückung wider, was zwar Vorteile in Bezug auf Formunterscheidung oder Diagnosesicherheit bieten kann, aber in Bezug auf die reine Erkennbarkeit nicht besser ist.

Ergebnisse der Bildqualität

Bildrauschen

Wie erwartet nahm das Rauschen (Standardabweichung in HU) mit steigender Strahlendosis, höheren iterativen Rekonstruktionsstufen und glatteren Precise Image Rekonstruktionsstufen ab (**Abbildung 2** und **Abbildung 3**). Im Vergleich zur gefilterten Rückprojektion und iterativen Rekonstruktion war das Rauschen bei Precise Image um bis zu 81% bzw. 73% reduziert (**Tabelle 1** und **Tabelle 2**). Bei allen Precise Image Stufen führte eine Erhöhung der Strahlendosis von 0,5 mGy auf 10 mGy zu einer durchschnittlichen Verringerung des Rauschens um das etwa 3,7-Fache.

Die mittlere Raumfrequenz des Rauschens nahm bei der Rekonstruktionsstufe „Sharper“ von Precise Image zu (**Abbildung 4** und **Abbildung 5**). Bei jeder Stufe war die mittlere Raumfrequenz konstant (<3% Abweichung) für alle Strahlendosen für die Stufen „Standard“, „Sharp“ und „Sharper“ (**Tabelle 1** und **Tabelle 2**). Für die Stufen „Smooth“ und „Smoother“ wurde eine maximale Abweichung von 12% festgestellt. Die mittleren Raumfrequenzen des Rauschleistungsspektrums waren für alle Precise Image Stufen bei der Einstellung für Lunge höher als bei der Einstellung für Weichgewebe.

Abbildung 2: Rauschen von Precise Image für alle Rekonstruktionsstufen und Strahlendosen für die beiden Aufgaben mit dem niedrigsten Kontrast, basierend auf der Einstellung für Weichgewebe

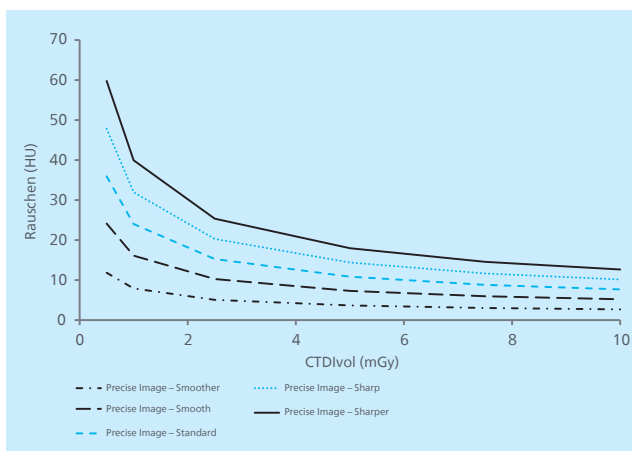


Abbildung 3: Rauschen von Precise Image für alle Rekonstruktionsstufen und Strahlendosen für die Aufgabe mit hohem Kontrast, basierend auf der Einstellung für Lunge

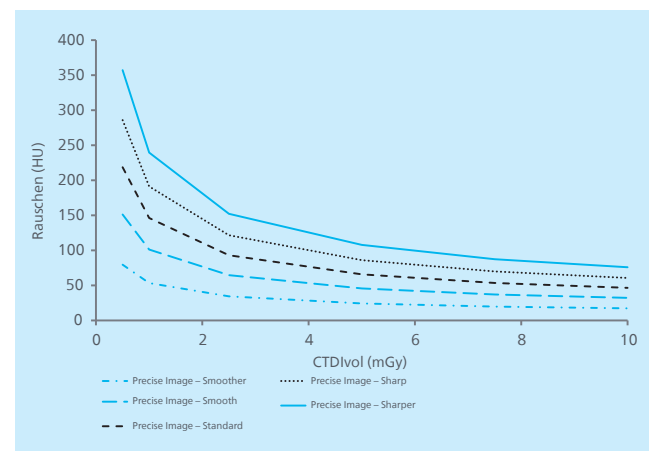


Abbildung 4: Mittlere Raumfrequenz (f_{mittel}) des Rauschens von Precise Image für alle Rekonstruktionsstufen und Strahlendosen für die beiden Aufgaben mit dem niedrigsten Kontrast, basierend auf der Einstellung für Weichgewebe

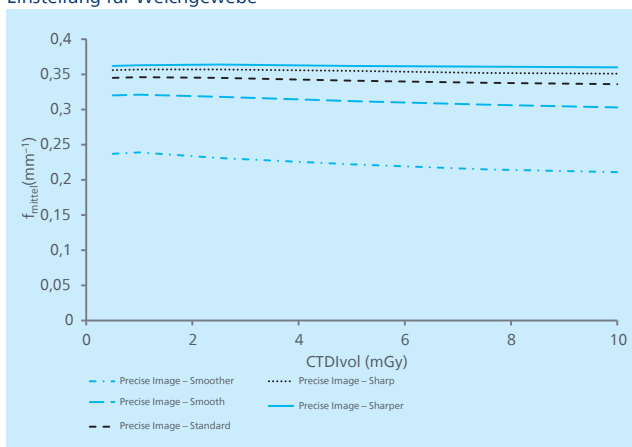
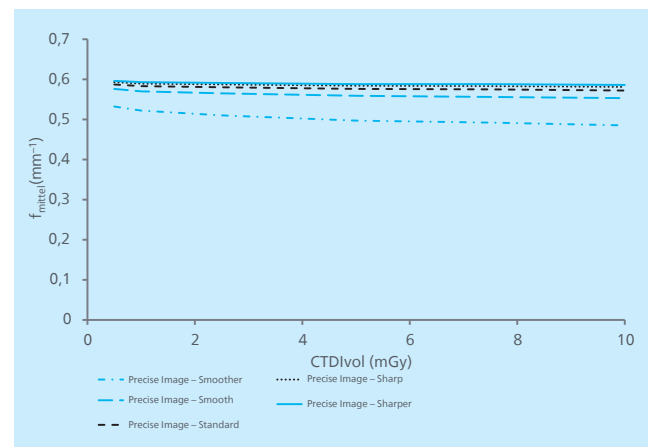


Abbildung 5: Mittlere Raumfrequenz (f_{mittel}) des Rauschens von Precise Image für alle Rekonstruktionsstufen und Strahlendosen für die Aufgabe mit hohem Kontrast, basierend auf der Einstellung für Lunge



	CTDI _{vol} (mGy)	Gefilterte Rück- projektion (FBP)	Iterative Rekonstruktionsstufe		Precise Image				
			4	7	Smoother	Smooth	Standard	Sharp	Sharper
Rauschen (HU)	0,5	61,8	44,5	30,1	11,8	24,1	35,9	47,8	59,8
	1	41,3	29,7	20,1	7,9	16,1	24,0	32,0	39,9
	2,5	26,1	18,8	12,7	5,1	10,3	15,3	20,3	25,4
	5	18,4	13,3	9,0	3,7	7,3	10,8	14,4	18,0
	7,5	14,9	10,8	7,3	3,0	6,0	8,8	11,7	14,6
	10	12,9	9,3	6,4	2,7	5,2	7,7	10,2	12,7
Mittlere Raum- frequenz des Rauschleistungs- spektrums (mm ⁻¹)	0,5	0,318	0,292	0,248	0,237	0,320	0,345	0,356	0,362
	1	0,321	0,294	0,250	0,239	0,321	0,346	0,357	0,363
	2,5	0,322	0,295	0,250	0,231	0,318	0,345	0,357	0,364
	5	0,321	0,293	0,247	0,222	0,312	0,341	0,355	0,362
	7,5	0,320	0,292	0,245	0,215	0,307	0,338	0,352	0,361
	10	0,320	0,290	0,242	0,211	0,303	0,336	0,351	0,360

Tabelle 1: Rauschen (Hounsfield-Einheiten [HU]) und mittlere Raumfrequenzen des Rauschleistungsspektrums für alle Strahlendosen und Rekonstruktionen für die beiden Aufgaben mit dem niedrigsten Kontrast (B-Filter für gefilterte Rückprojektion und iterative Rekonstruktion, Einstellung für Weichgewebe für Precise Image)

	CTDI _{vol} (mGy)	Gefilterte Rück- projektion (FBP)	Iterative Rekonstruktionsstufe		Precise Image				
			4	7	Smoother	Smooth	Standard	Sharp	Sharper
Rauschen (HU)	0,5	257,9	183,3	117,6	79,5	151,1	218,5	285,8	357,0
	1	172,6	122,2	77,7	53,3	101,1	146,1	191,1	239,5
	2,5	109,5	77,5	49,3	34,3	64,5	93,0	121,5	152,2
	5	77,5	54,8	34,8	24,2	45,6	65,8	86,0	107,7
	7,5	62,9	44,5	28,3	19,7	37,0	53,4	69,8	87,4
	10	54,6	38,6	24,5	17,2	32,2	46,4	60,6	75,9
Mittlere Raum- frequenz des Rauschleistungs- spektrums (mm ⁻¹)	0,5	0,523	0,519	0,515	0,532	0,576	0,587	0,593	0,596
	1	0,522	0,518	0,513	0,522	0,570	0,583	0,589	0,593
	2,5	0,522	0,518	0,513	0,510	0,565	0,580	0,587	0,591
	5	0,520	0,516	0,512	0,497	0,559	0,576	0,584	0,588
	7,5	0,519	0,515	0,511	0,492	0,556	0,575	0,583	0,588
	10	0,519	0,515	0,510	0,485	0,553	0,572	0,581	0,586

Tabelle 2: Rauschen (Hounsfield-Einheiten [HU]) und mittlere Raumfrequenzen des Rauschleistungsspektrums für alle Strahlendosen und Rekonstruktionen für die Aufgabe mit hohem Kontrast (YA-Filter für gefilterte Rückprojektion und iterative Rekonstruktion, Einstellung für Lunge für Precise Image)

Bildauflösung: aufgabenbasierte Übertragungsfunktion

Im Vergleich zur gefilterten Rückprojektion und iterativen Rekonstruktion war die Bildauflösung bei allen Precise Image Stufen mit erhöhten $TTF_{50\%}$ -Werten für den Lufteinsatz erhöht (**Tabelle 3**). Für den Polyethyleinsatz zeigten sich gleiche Ergebnisse für alle Dosen über 0,5 mGy.

	CTDI _{vol} (mGy)	Gefilterte Rück- projektion (FBP)	Iterative Rekonstruktionsstufe		Precise Image				
			4	7	Smoother	Smooth	Standard	Sharp	Sharper
TTF_{50%} – Luft (mm⁻¹)	0,5	0,499	0,496	0,492	0,600	0,606	0,643	0,768	0,780
	1	0,586	0,483	0,496	0,715	0,729	0,603	0,595	0,692
	2,5	0,501	0,514	0,527	0,680	0,696	0,688	0,673	0,659
	5	0,557	0,578	0,602	0,687	0,773	0,735	0,802	0,789
	7,5	0,574	0,500	0,511	0,728	0,719	0,723	0,717	0,717
	10	0,584	0,597	0,536	0,701	0,704	0,695	0,685	0,742
TTF_{50%} – Polyethylen (mm⁻¹)	0,5	0,254	0,238	0,218	0,209	0,212	0,228	0,173	0,212
	1	0,314	0,283	0,257	0,329	0,332	0,351	0,354	0,357
	2,5	0,306	0,268	0,294	0,323	0,327	0,327	0,369	0,356
	5	0,319	0,300	0,280	0,381	0,369	0,376	0,369	0,373
	7,5	0,317	0,304	0,297	0,363	0,368	0,365	0,362	0,365
	10	0,327	0,319	0,287	0,380	0,380	0,375	0,375	0,373

Tabelle 3: Aufgabenbasierte Übertragungsfunktion bei 50% (TTF50%) für den Luft- und Polyethyleinsatz für gefilterte Rückprojektion, iterative Rekonstruktion und alle Precise Image Stufen für alle Strahlendosen. Für den Lufteinsatz basieren die Ergebnisse auf dem YA-Filter für gefilterte Rückprojektion und iterative Rekonstruktion sowie auf der Einstellung für Lunge für Precise Image. Für den Polyethyleinsatz basieren die Ergebnisse auf dem B-Filter für gefilterte Rückprojektion und iterative Rekonstruktion sowie auf der Einstellung für Weichgewebe für Precise Image.

Erkennbarkeitsindex

Bei Precise Image stieg der Erkennbarkeitsindex für alle simulierten Läsionen mit zunehmender Glättung und Strahlendosis an (**Abbildung 6**, **Abbildung 7** und **Abbildung 8**). Bei den beiden Aufgaben mit dem niedrigsten Kontrast war der Erkennbarkeitsindex bei den Precise Image Rekonstruktionen höher als bei der klassischen gefilterten Rückprojektion und iterativen Rekonstruktion (**Abbildung 6** und **Abbildung 7**). Die Stufe „Smoother“ lieferte den höchsten Erkennbarkeitsindex, unabhängig von der Strahlendosis. Bei der Aufgabe mit hohem Kontrast und inhärent hohen Erkennungsrate verbesserte sich der Erkennbarkeitsindex bei der Precise Image Rekonstruktion nicht (**Abbildung 8**).

Abbildung 6: Erkennbarkeitsindex (d') für gefilterte Rückprojektion, iterative Rekonstruktion (iDose Stufe 4 und 7) und alle verfügbaren Precise Image Stufen für die Erkennung des Thorax-Rundherds mit geringem Kontrast im Weichgewebe (Kontrast 50 Hounsfield-Einheiten [HU]) auf der Grundlage des Rauschleistungsspektrums und der aufgabenbasierten Übertragungsfunktion mit Polyethyleinsatz, beide mit der Einstellung für Weichgewebe

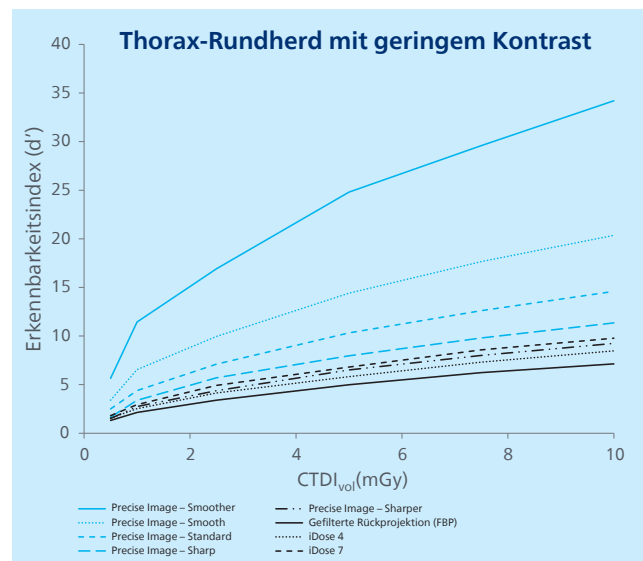


Abbildung 7: Erkennbarkeitsindex (d') für gefilterte Rückprojektion, iterative Rekonstruktion (iDose Stufe 4 und 7) und alle verfügbaren Precise Image Stufen für die Erkennung der milchglasartigen Dichteanhebung (Kontrast 200 HU) auf der Grundlage des Rauschleistungsspektrums und der aufgabenbasierten Übertragungsfunktion mit Polyethyleinsatz, beide mit der Einstellung für Weichgewebe

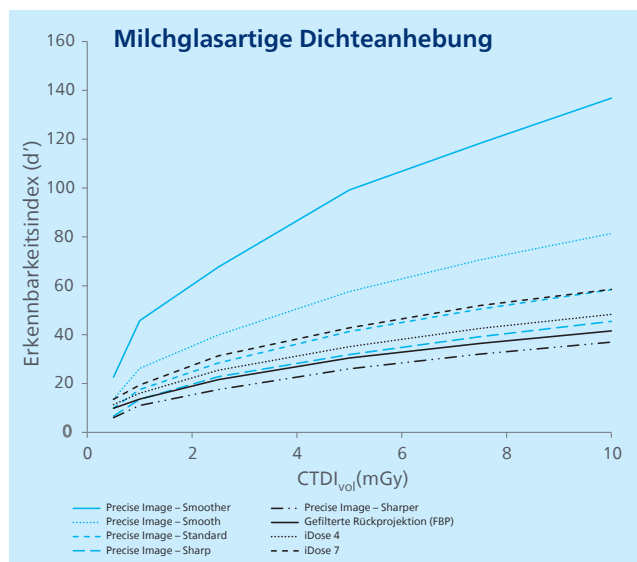
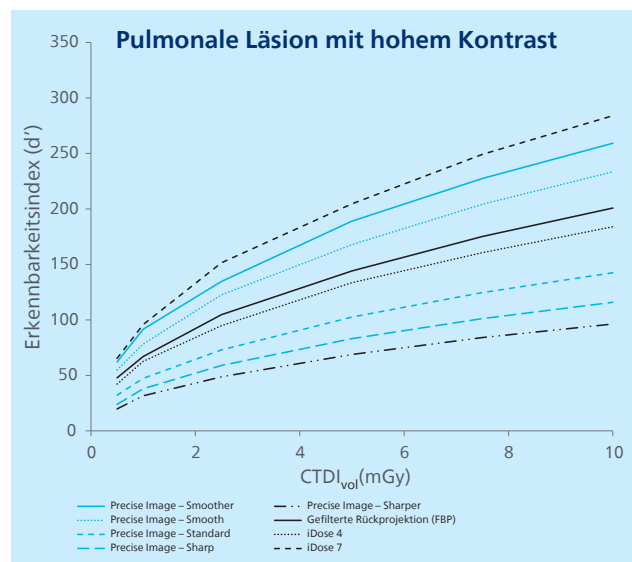


Abbildung 8: Erkennbarkeitsindex (d') für gefilterte Rückprojektion, iterative Rekonstruktion (iDose Stufe 4 und 7) und alle verfügbaren Precise Image Stufen für die Erkennung der pulmonalen Läsion mit hohem Kontrast (50 HU) auf der Grundlage des Rauschleistungsspektrums und der aufgabenbasierten Übertragungsfunktion mit Lufteinsatz, beide mit der Einstellung für Lunge



Vergleich verschiedener Anbieter

Im Vergleich zu der zuvor veröffentlichten Analyse zur DLR von zwei anderen Anbietern war das Rauschen von Precise Image mit der Einstellung für Weichgewebe und den Stufen „Smoother“, „Smooth“ und „Standard“ vergleichbar.³ Die Rekonstruktionsstufen „Sharp“ und „Sharper“ sowie alle auf der Einstellung für Lunge basierenden Stufen führten zu erhöhtem Bildrauschen. Bei der Einstellung für Weichgewebe waren die mittleren Raumfrequenzen des Rauschleistungsspektrums ebenfalls mit der zuvor veröffentlichten DLR-Analyse vergleichbar. Für die Einstellung für Lunge zeigten sich für Precise Image erhöhte mittlere Raumfrequenzen des Rauschleistungsspektrums.

Beim Polyethyleinsatz war die Bildauflösung (basierend auf der TTF bei 50%) für Precise Image vergleichbar mit den zuvor berichteten Werten für die anderen Anbieter.³ Beim Lufteinsatz zeigte Precise Image eine höhere Bildauflösung, insbesondere für die schärferen Rekonstruktionsstufen.

Die d' -Werte von Precise Image (Stufe „Smoother“) waren, unabhängig von der Strahlendosis, höher als die der beiden anderen DLR-Rekonstruktionstechniken (**Abbildung 9** und **Abbildung 10**). Für Dosen von 1 mGy und höher galt dies sowohl für die Rekonstruktionsstufe „Smoother“ als auch „Smooth“. Für die pulmonale Läsion mit hohem Kontrast und mit inhärent hohem Bildkontrast wurde der höchste Erkennbarkeitsindex bei AiCE erzielt, unabhängig von der Strahlendosis (**Abbildung 11**).

Abbildung 9: Erkennbarkeitsindex (d') für alle verfügbaren Deep-Learning-Rekonstruktionsalgorithmen für die Erkennung des Thorax-Rundherds mit geringem Kontrast im Weichgewebe (Kontrast 50 HU)

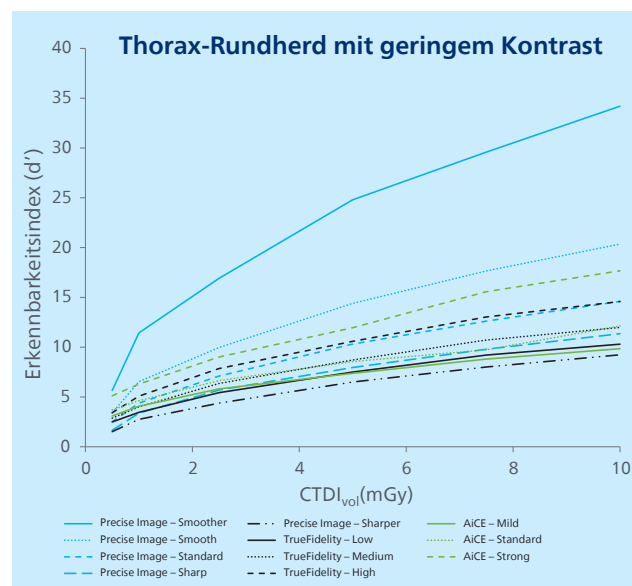


Abbildung 10: Erkennbarkeitsindex (d') für alle verfügbaren Deep-Learning-Rekonstruktionsalgorithmen für die Erkennung der milchglasartigen Dichteanhebung (Kontrast 200 HU)

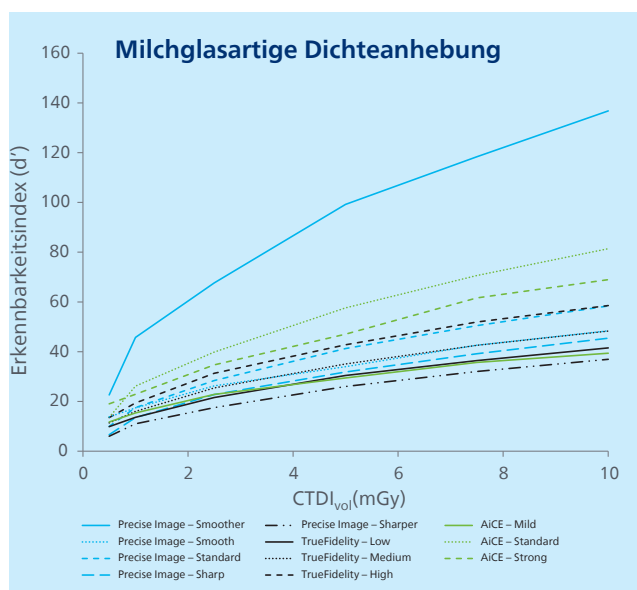
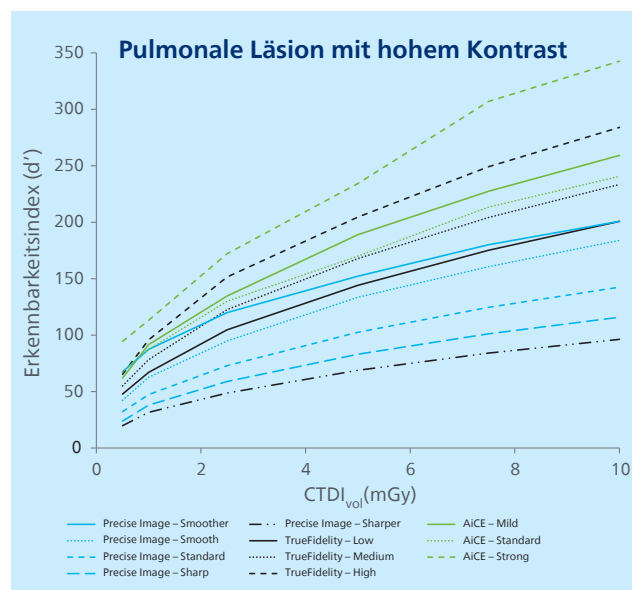


Abbildung 11: Erkennbarkeitsindex (d') für alle verfügbaren Deep-Learning-Rekonstruktionsalgorithmen für die Erkennung der pulmonalen Läsion mit hohem Kontrast (950 HU)



Fazit

Precise Image verbesserte die Erkennbarkeitsindizes für die drei untersuchten simulierten Läsionen im Vergleich zur klassischen gefilterten Rückprojektion und iterativen Rekonstruktion signifikant. Außerdem zeigte Precise Image im Vergleich zu den Mitbewerbern deutlich bessere Erkennbarkeitsindizes für Thorax-Rundherde mit geringem Kontrast und milchglasartige Dichteanhebungen. Bei den pulmonalen Läsionen mit hohem Kontrast war die Bildqualität bei allen Anbietern vergleichbar, wobei der Erkennbarkeitsindex insgesamt hoch war.

1. Willemink MJ, Noël PB. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. *Eur Radiol.* 2019;29:2185–2195.
2. Smith TB, Solomon J, Samei E. Estimating detectability index in vivo: development and validation of an automated methodology. *Journal of Medical Imaging.* 2017;5:1.
3. Greffier J, Frandon J, Si-Mohamed S, et al. Comparison of two deep learning image reconstruction algorithms in chest CT images: A task-based image quality assessment on phantom data. *Diagn Interv Imaging.* 2022;103:21–30.
4. Greffier J, Si-Mohamed S, Frandon J, et al. Impact of an artificial intelligence deep-learning reconstruction algorithm for CT on image quality and potential dose reduction: A phantom study. *Med Phys.* 2022;49:5052–5063.
5. Greffier J, Boccalini S, Beregi JP, et al. CT dose optimization for the detection of pulmonary arteriovenous malformation (PAVM): A phantom study. *Diagn Interv Imaging.* 2020;101:289–297.
6. Greffier J, Hamard A, Pereira F, et al. Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. *Eur Radiol.* 2020;30:3951–3959.
7. Solomon J, Samei E. Correlation between human detection accuracy and observer model-based image quality metrics in computed tomography. *J Med Imag.* 2016;3:35506.
8. Barrett HH, Myers KJ. *Foundations of Image Science.* Hoboken, NJ: John Wiley & Sons, 2004.

