

Aus dem Charité Centrum für  
diagnostische und interventionelle Radiologie und Nuklearmedizin  
Klinik für Radiologie (mit dem Bereich Kinderradiologie)

DISSERTATION

Vergleich zwischen homogenen und realistischen Phantomen zur  
Bewertung von Bildqualität in der Computertomographie –  
Phantomeinfluss auf Niedrigkontrastdetektierbarkeit

-

Comparison of uniform and realistic phantoms for image quality  
evaluation in computed tomography – phantom effects on low-contrast  
detectability

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Juliane Franziska Conzelmann

Erstbetreuer: PD Dr. Paul Jahnke

Datum der Promotion: 25.06.2023



## Inhaltsverzeichnis

Tabellenverzeichnis.....	iii
Abbildungsverzeichnis.....	iv
Abkürzungsverzeichnis.....	v
Zusammenfassung .....	1
1 Einleitung.....	3
1.1 Überblick über den aktuellen Forschungsstand.....	3
1.2 Fragestellungen.....	9
2 Methodik.....	10
2.1 In dieser Arbeit verwendete Phantome.....	10
2.2 CT-Aufnahmen .....	11
2.3 Auswahl der CT-Bilder.....	11
2.4 Aufbau des Detektionsexperimentes .....	11
2.5 Charakterisierung des Rauschens und Noise-Power-Spektrums.....	13
2.6 Datenanalyse .....	13
3 Ergebnisse .....	14
3.1 Detektion von Niedrigkontrastläsionen im Vergleich zwischen homogenen und anatomischen Phantomen .....	14
3.2 Einfluss der Dosis und Rekonstruktionsmethode auf die Detektion von Niedrigkontrastläsionen.....	15
3.3 Rauscheigenschaften und Noise-Power-Spektrum .....	17
4 Diskussion .....	19
4.1 Allgemeine Zielstellung.....	19
4.2 Zusammenfassung der wichtigsten Ergebnisse und Einordnung in die Literatur .....	20
4.3 Limitationen der Studie.....	22
4.4 Implikationen für Praxis und/oder zukünftige Forschung .....	22

---

5	Schlussfolgerungen .....	23
	Literaturverzeichnis .....	24
	Eidesstattliche Versicherung .....	30
	Anteilerklärung an den erfolgten Publikationen.....	31
	Auszug aus der Journal Summary List .....	33
	Druckexemplar(e) der Publikation(en) .....	34
	Lebenslauf .....	43
	Komplette Publikationsliste.....	44
	Danksagung .....	45

## Tabellenverzeichnis

- Tabelle 1:** Detektionsraten (Prozent  $\pm$  Standardabweichung) im homogenen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: Modifiziert nach Conzelmann, J et al. 2021. .... 15
- Tabelle 2:** Detektionsraten (Prozent  $\pm$  Standardabweichung) im anatomischen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: modifiziert nach Conzelmann, J et al. 2021. .... 15
- Tabelle 3:** Diagnostische Zuversicht (Mittelwert  $\pm$  Standardabweichung) im homogenen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: Eigene Darstellung: Conzelmann, J. .... 16
- Tabelle 4:** Diagnostische Zuversicht (Mittelwert  $\pm$  Standardabweichung) im anatomischen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: Eigene Darstellung: Conzelmann, J. .... 16
- Tabelle 5:** Rauschwerte als Mittelwert  $\pm$  Standardabweichung der gemessenen Pixelstandardabweichungen. Quelle: Modifiziert nach Conzelmann, J et al. 2021 ..... 18
- Tabelle 6:** Räumliche Frequenz des Noise Power Spektrum Maximums ( $\text{mm}^{-1}$ ). Quelle: Modifiziert nach Conzelmann, J. et al 2021 ..... 18

## Abbildungsverzeichnis

**Abbildung 1:** Photographien der in dieser Arbeit verwendeten Phantome: Homogenes Phantom aus Polymethylmethacrylat (links) und Beispiel eines 3D-gedruckten anatomischen Phantoms (rechts). Quelle: Eigenes Bildmaterial. .... 10

**Abbildung 2:** Aufbau des 4-AFC-Experiments. Beispiel für Bildquartette des anatomischen Phantoms (links) und des homogenen Phantoms (rechts). Über den Quartetten wird jeweils ein Referenzbild angezeigt, das die Position der zu detektierenden Läsion angibt. Quelle: Eigenes Bildmaterial..... 12

## Abkürzungsverzeichnis

<b>3D</b>	Dreidimensional
<b>4-AFC</b>	4-Alternative-forced-choice
<b>AIDR 3D</b>	Englisch: Adaptive iterative dose reduction 3D, Deutsch: Adaptive iterative Dosisreduktion 3D
<b>CDTIvol</b>	Computed tomography dose index
<b>CNR</b>	Englisch: Contrast-to-noise-ratio, Deutsch: Kontrast-zu-Rausch-Verhältnis
<b>CT</b>	Englisch: Computed tomography, Deutsch: Computertomographie
<b>FBP</b>	Englisch: Filtered back projection, Deutsch: Gefilterte Rückprojektion
<b>HE</b>	Hounsfield Einheit
<b>IR</b>	Iterative Rekonstruktion
<b>NPS</b>	Noise-Power-Spektrum
<b>ROI</b>	Region of interest
<b>SNR</b>	Englisch: Signal-to-noise-ratio, Deutsch: Signal-zu-Rausch-Verhältnis

## Zusammenfassung

Die Computertomographie (CT) spielt eine Schlüsselrolle bei der Diagnose und Verlaufskontrolle zahlreicher Erkrankungen, geht allerdings auch mit der Strahlenexposition von Patienten und Patientinnen einher. Ein Hauptziel moderner CT-Technologien und klinischer Bildgebungsprotokolle besteht daher darin, ein optimales Gleichgewicht zwischen Strahlenexposition und angemessener Bildqualität für eine zuverlässige Diagnose zu erreichen.

Phantome sind Referenzobjekte, die zu Testzwecken untersucht werden, um CT-Technologien und Bildgebungsverfahren zu überprüfen und zu verbessern. Phantome ermöglichen wiederholte CT-Aufnahmen, vermeiden die Strahlenexposition von Patienten und Patientinnen für Testzwecke und liefern eine bekannte Grundwahrheit zum Abgleich mit dem Scanergebnis. Derzeit besitzen Phantome häufig homogene Hintergrundstrukturen, wohingegen Patienten und Patientinnen anatomische Details und unterschiedliche Gewebestrukturen aufweisen. Es ist daher fraglich, inwieweit die Bewertung von CT-Aufnahmen mit homogenen Phantomen auf die klinische Bildgebung übertragen werden kann.

Vor diesem Hintergrund wurde in der vorliegenden Arbeit der Einfluss der Phantomhintergrundstruktur auf die Bewertung von CT-Bildqualität in Niedrigkontrastdetektierbarkeitsexperimenten untersucht. Hierfür wurden zwei Halsphantome mit identischer Form und Größe verwendet, die sich durch ihre homogenen beziehungsweise anatomischen Hintergrundstrukturen unterschieden. Sieben Radiologen und Radiologinnen bewerteten die Detektierbarkeit von Niedrigkontrastläsionen. Darüber hinaus wurden Rauschparameter zwischen den beiden Phantomarten verglichen.

Die Ergebnisse zeigten, dass die Detektion von Niedrigkontrastläsionen bei gleichem Läsionskontrast zu umliegenden Arealen in beiden Phantomarten signifikant durch anatomische Hintergrundstrukturen beeinträchtigt wird. Darüber hinaus wurde die Bewertung von Dosis und Bildrekonstruktionsverfahren durch Phantomeigenschaften beeinflusst und die Rauschunterdrückung durch iterative Rekonstruktion war abhängig von den im Phantom vorliegenden Hintergrundstrukturen.

Zusammenfassend ergibt sich, dass Hintergrundstrukturen und anatomische Details von Phantomen die Bewertung von diagnostischer Bildqualität und Bildgebungstechnologien



beeinflussen. Es ist davon auszugehen, dass die Aussagekraft von CT-Bewertungen für die klinische Bildgebung mit der Realitätsnähe der Untersuchungsumgebung zunimmt.

## **Abstract**

Computed tomography (CT) plays a key role for the diagnosis and follow up of many diseases but also involves the exposure of patients to ionizing radiation. A main goal of modern CT technologies and clinical imaging protocols is therefore to achieve an optimal balance between dose exposure and adequate image quality for a reliable diagnosis.

Phantoms are reference objects that are scanned for test purposes to verify and improve CT technologies and imaging methods. Phantoms enable repeated CT scans without exposing patients to radiation and provide a ground truth for comparison with the scanner output. Phantoms frequently have a uniform background structure, whereas patients have anatomical details and tissue structures. It is therefore questionable to which extent CT assessment with uniform phantoms can be transferred to clinical imaging.

This work investigated the influence of phantom background structure on CT assessment using the detectability of low-contrast lesions as an image quality parameter. Two neck phantoms of identical shape and with uniform and anatomical background structures were examined. Low-contrast detectability was assessed by seven radiologists. Additionally, noise metrics were compared between both phantom types.

Results showed that lesion detection was significantly impaired by anatomical background structure in a comparison of lesions with identical contrast to surrounding structures. Moreover, assessment of dose and image reconstruction was influenced by the phantom type and the denoising power of iterative reconstruction was dependent on phantom background structure.

In conclusion, phantom background structure and anatomical detail influence the evaluation of diagnostic image quality and imaging technologies. The validity of CT assessment for clinical imaging can be expected to increase as the experimental design becomes more realistic.

# 1 Einleitung

## 1.1 Überblick über den aktuellen Forschungsstand

Die Computertomographie (CT) ist eines der wichtigsten bildgebenden Verfahren in der radiologischen Diagnostik, geht allerdings auch immer mit der Strahlenexposition von Patienten und Patientinnen einher. Der Einsatz der CT als schnelle und kostengünstige Methode hat in den vergangenen Jahren beständig zugenommen und durch technologische Entwicklungen neuer Anwendungsfelder erschlossen. Die Bedeutung der CT für die moderne Diagnostik wird durch jährlich ca. 13 Millionen Untersuchungen verdeutlicht, die allein in Deutschland durchgeführt werden (1).

Die Strahlendosis konnte seit Einführung der CT in den 1970er Jahren durch die Optimierung von Aufnahmeprotokollen und Anpassung von Scanparametern an die Patientenanatomie sukzessive gesenkt werden (2,3). Im Fokus standen Entwicklungen der Gerätetechnik und Bildverarbeitung, die zu einer effizienteren Nutzung von Strahlung, besserer Bildqualität und Reduzierung der Strahlenexposition führten. Dem gegenüber steht der zahlenmäßige Anstieg an Untersuchungen und der damit verbundenen kumulativen Strahlenbelastung der Bevölkerung (1). Die CT verwendet ionisierende Strahlung, die zu DNA-Schäden führt und mutagenes Risiko beinhaltet (4). Dies gilt grundlegend für alle Patienten und Patientinnen, wobei das Krebsrisiko durch CT-Untersuchungen bei jüngeren Patienten und Patientinnen besonders hoch ist (5).

Daher liegt der Strahlenexposition in der CT das „ALARA“-Prinzip („as low as reasonably achievable“) zugrunde. Demnach sollte stets die geringste Strahlendosis verwendet werden, die notwendig ist, um eine für die Diagnosestellung ausreichende Bildqualität zu erreichen (6). Dabei hängen die Anforderungen an die Bildqualität von der diagnostischen Fragestellung ab. Beispielsweise besteht das primäre Ziel von Nierenstein-CT-Untersuchungen in der Darstellung von hochkontrastierten und dadurch verhältnismäßig einfach zu detektierenden Konkrementen, während bei Fragestellungen zu hepatischen oder zerebralen Metastasen häufig subtile Weichteilbefunde zu bewerten sind, wofür meist höhere Anforderungen an Dosis und Bildqualität gelten.

Zur objektiven Beurteilung von Bildqualität stehen verschiedene physikalische Messparameter zur Verfügung, mit denen sich Bildrauschen, Kontrast und Ortsauflösung beschreiben lassen (7–10).

Das Bildrauschen entsteht im Zuge des Bildgebungsprozesses als Überlagerung des realen Bildsignals und kann als Standardabweichung der Pixelwerte im Bild gemessen werden (11). Zu den Ursachen des Bildrauschens gehören das Quantenrauschen, das durch die stochastische Verteilung von Photonen auf dem Detektor entsteht, und das elektronische Rauschen, das Schwankungen der Signalweiterleitung aus dem Detektor widerspiegelt. Es ist darüber hinaus üblich, dass auch Textureinflüsse von Gewebe, die zwar nicht durch den Bildgebungsprozess selbst entstehen, aber dennoch die Bildauswertung erschweren, als anatomisches Rauschen dem Bildrauschen zugeordnet werden (11,12). Bildrauschen kann dazu führen, dass reale Signale, beispielsweise von kleinen Strukturen oder Läsionen, nicht wahrgenommen werden können und die diagnostische Aussagekraft von CT-Bildern abnimmt. Ein wesentliches Ziel zahlreicher CT-Technologien besteht daher darin, das Bildrauschen zu reduzieren.

Der Bildkontrast beschreibt den Signalunterschied von Strukturen und spielt beispielsweise eine Rolle, um benachbarte Gewebetypen voneinander zu unterscheiden (13). Durch Unterschiede in der Gewebedichte und -zusammensetzung und den darauf beruhenden linearen Schwächungskoeffizienten von Gewebe werden Röntgenstrahlen unterschiedlich stark abgeschwächt. Dies kommt in CT-Bildern durch verschiedene Graustufen zum Ausdruck und lässt sich in Hounsfield-Einheiten (HE) messen (14). HE sind ein standardisiertes Maß zur Beschreibung der Strahlenschwächung von Gewebe in CT-Bildern und werden in Bezug auf Wasser (0 HE) und Luft (-1000 HE) berechnet. Liegen die HE-Werte und Graustufen von benachbarten Geweben nah beieinander, besteht ein geringer Bildkontrast, der die Detektion von Befunden erschweren kann. Die leistungsstarke Darstellung von Kontrastunterschieden durch CT-Systeme spielt daher eine wichtige Rolle in der Bewertung von diagnostischer Bildqualität (13).

Um Bildrauschen und -kontrast einfach verständlich zusammenzufassen, hat es sich als hilfreich erwiesen, Kontrast und Rauschen ins Verhältnis zu setzen und als Kontrast-zu-Rausch-Verhältnis (englisch contrast-to-noise ratio, CNR) zu beschreiben (12). Mit steigendem CNR, also hohen Kontrast- und niedrigen Rauschwerten, wird die Bildqualität durch Radiologen und Radiologinnen zunehmend positiv bewertet (15,16). Das CNR hängt wiederum von Aufnahmeparametern wie Röhrenstrom und Röhrenspannung ab. Höhere Röhrenströme reduzieren das Bildrauschen und steigern dadurch das CNR. Allerdings gehen höhere Röhrenströme auch mit höheren Dosen einher. Die Röhrenspannung bestimmt das Energiespektrum der Röntgenstrahlung und beeinflusst dadurch sowohl den Bildkontrast als auch das Rauschen. Niedrigere Röhrenspannungen können

eingesetzt werden, um die Dosis zu reduzieren und den Kontrast zu erhöhen, führen allerdings auch zu vermehrtem Bildrauschen.

Darüber hinaus haben Rekonstruktionsverfahren einen wesentlichen Einfluss auf das Bildrauschen und das Kontrast-zu-Rausch-Verhältnis. Rekonstruktionsverfahren werden benötigt, um die aufgezeichneten Rohdaten der Strahlungsschwächungswerte in ein digitales CT-Bild zu konvertieren. Traditionell wurden CT-Bilder mittels gefilterter Rückprojektion (englisch: filtered back projection, FBP) rekonstruiert. Bei der FBP wird durch die Rückprojektion der Rohdaten an ihren ursprünglichen Ort ein zweidimensionales Bild kreiert. Dieses Bild ist noch unscharf und wird mit Hilfe eines Filters (Kernels) geschärft, bis das fertige CT-Bild entsteht (17). Die FBP ist ein lineares Rekonstruktionsverfahren, bei dem die Dosis mit dem Bildrauschen korreliert. Ein zentrales Ziel in der Weiterentwicklung der Bildrekonstruktionsverfahren der vergangenen Jahre bestand daher darin, das Rauschen im Zuge der Bildverarbeitung zu reduzieren, um auch bei niedriger Dosis ein angemessenes CNR zu erzielen (18–20). Als besonders effektiv haben sich hierbei iterative Rekonstruktionsverfahren (IR) erwiesen. Bei dieser Rekonstruktionstechnik werden zur Bilderzeugung Korrekturschleifen eingebaut. Ein CT-Bild des Objektes wird mit FBP erstellt und wiederholt mit den tatsächlich gemessenen Strahlungsschwächungswerten verglichen und korrigiert, bis eine ausreichende Übereinstimmung erzielt wird (21). Die IR ermöglicht Bildaufnahmen mit geringeren Dosen, ohne dass das Bildrauschen wesentlich ansteigt. In diesem Zusammenhang wurde über ein Dosisreduktionspotential von bis zu 70% im Vergleich zur konventionellen FBP berichtet (22). Die IR hat sich dadurch bei modernen CT-Scannern als Standardverfahren zur Bildrekonstruktion durchgesetzt.

Die IR unterdrückt das Bildrauschen effektiv, allerdings kommt es vor allem bei niedrigen Strahlendosen zu Veränderungen der Bildtextur. Diese werden durch Bildqualitätsparameter wie Bildrauschen und CNR nicht erfasst, können sich aber durch das Noise-Power-Spektrum (NPS) charakterisieren lassen (7,12). Das NPS beruht auf der Fourier-Transformation von CT-Bildern und beschreibt die Textur des Rauschens anhand der Amplitude in unterschiedlichen Frequenzbereichen. Es ist somit Maß für verschiedene Rauschtexturen, die zu einer grobkörnigeren (überwiegend niedrige Frequenzen) oder feinkörnigeren (überwiegend höhere Frequenzen) Textur führen kann. IR führt insbesondere bei niedrigen Dosen zu einer gröberen Textur, die sich im NPS durch Verschiebungen zu niedrigeren Frequenzen widerspiegelt und Befunde verzerren oder unkenntlich machen kann. Das NPS hat allerdings Einschränkungen, da die Verwendung des NPS

unter der Annahme stattfindet, dass sich das Rauschen gleichmäßig über das Bild verteilt. Diese trifft bei der iterativen Rekonstruktion jedoch bekanntermaßen nicht zu (7). Um dennoch möglichst aussagekräftige Ergebnisse zu erhalten, wird das NPS in homogenen Bildregionen bestimmt, unter der Annahme, dass das Rauschen sich dort am gleichmäßigsten verteilt. Somit ist bei der Interpretation zu beachten, dass das NPS keine Auskunft über die Rauschtextur im gesamten Bild liefert, sondern lediglich die Rauschtextur in einem begrenzten Ausschnitt widerspiegelt (23).

Da weder Bildrauschen noch CNR oder NPS eine umfassende und aussagekräftige Beschreibung der Bildqualität liefern, werden objektive Methoden benötigt, um die diagnostische Leistung bildgebender Systeme zu beurteilen. Dies gilt insbesondere im Zusammenhang mit nicht-linearen Rekonstruktionsverfahren wie IR (7). In Anlehnung an die diagnostischen Prozesse in der klinischen Routine werden hierfür häufig Experimente zur Ermittlung der Detektionsrate von niedrigkontrastierten Signalen als Bewertungskriterium verwendet. Hierbei werden CT-Bilder sowohl mit als auch ohne niedrigkontrastierte Läsionen vorgelegt und durch Studienteilnehmende oder mathematische Algorithmen ausgewertet. Voraussetzung für solche Bewertungen ist, dass Kenntnis darüber besteht, ob in den präsentierten Bildern tatsächlich Läsionen abgebildet wurden. Dies erfordert in der Regel die Verwendung von Phantomen, die darüber hinaus keine Einschränkungen in der Exposition gegenüber ionisierender Röntgenstrahlung aufweisen. Dies ist vorteilhaft, da systematische Untersuchungen von CT-Technologien häufig umfangreiche CT-Aufnahmen erfordern, die auf Grund der Strahlenrisiken mit Patienten und Patientinnen nicht durchzuführen wären (4).

Bisher wurden dafür überwiegend einfache geometrische Phantome verwendet, die sich aus homogenen Materialien zusammensetzten. Beliebte Phantome sind beispielsweise das ACR-Phantom (Gammex) und das Catphan-Phantom (Phantom Laboratories) (24). Diese Phantome enthalten Niedrigkontrastläsionen in unterschiedlichen Größen und mit verschiedenen Kontrasten. Beispielsweise enthält das Catphan-Phantom Läsionen mit 2 bis 15 mm Durchmesser und 0,3 bis 1% Kontrast (22). Eine deutliche Einschränkung der homogenen Phantome liegt vor allem in ihrer unnatürlichen homogenen Hintergrundstruktur, die die Komplexität der menschlichen Anatomie und Gewebedetails nicht berücksichtigt.

Bisher wurde in den meisten Untersuchungen davon ausgegangen, dass Ergebnisse, die mittels homogener Phantome erhoben wurden, auf die klinische Bildgebung und Diagnostik von Patienten und Patientinnen übertragbar sind. Aus vorangegangenen Röntgen-

und CT-Studien ist jedoch bekannt, dass anatomische Details und komplexe Hintergrundstrukturen die Bewertung von in Anlehnung an die klinische Diagnostik erhobenen Bildparametern wie die Detektionsrate von Läsionen beeinflussen (25–27). Es ist daher äußerst fragwürdig, inwiefern Bildqualitätsbewertungen mittels einfacher homogener Phantome tatsächlich auf die klinische Bildgebung von Patienten und Patientinnen übertragbar sind.

Es gibt verschiedene Ansätze, um die menschliche Anatomie in Detektionsstudien zu berücksichtigen. So wurden beispielsweise CT-Aufnahmen von Körperspenden verwendet (28). In einem anderen Ansatz wurden CT-Bilder digital nachbearbeitet, indem Läsionen eingefügt und verschiedene CT-Einstellungen durch Bildbearbeitung nachgeahmt wurden (29). Es gibt darüber hinaus auch Ansätze, anatomisch realistische Phantome zu entwickeln und zu verwenden. Hierzu gehören semianthropomorphe Phantome mit einer vereinfachten Darstellung anatomischer Strukturen wie z.B. der Leber (30) und anthropomorphe Phantome, die beispielsweise Lungengewebe abbilden (31). Die Nachbildung komplexer anatomischer Strukturen und die Anpassung von Phantomen an bestimmte Aufgabenstellungen war jedoch bislang nur eingeschränkt möglich.

3D-Drucktechnologien bieten in diesem Zusammenhang die Möglichkeit, ausgehend von realen CT-Datensätzen Phantome herzustellen, die die Anatomie und Pathologie des Menschen realistischer wiedergeben und für Bildbewertungen anhand diagnostischer Auswertungen verwendet werden können. Als besonders leistungsstark hat sich ein kombiniertes Verfahren aus Tintenstrahldruck und papierbasiertem 3D-Druck erweisen, das unter dem Begriff strahlendichter 3D-Druck zusammengefasst wird (32). Bei diesem Verfahren werden CT-Bilder zunächst mit iodhaltigen Tinten auf Papier gedruckt, wodurch anatomische und pathologische Details verlustfrei übertragen werden können. Anschließend werden die bedruckten Papiere gestapelt, verklebt und auf die Patientenform zugeschnitten, wodurch dreidimensionale Objekte mit der Form, Größe und den in dem als Vorlage verwendeten Datensatz hinterlegten Schwächungseigenschaften entstehen (33).

Das 3D-Druckverfahren ermöglicht dadurch die Herstellung von Phantomen, die individuelle Patienten und Patientinnen einschließlich anatomischer und pathologischer Details nachbilden. Darüber hinaus eröffnet das Verfahren die Möglichkeit, Datensätze vor der Herstellung so zu manipulieren, dass niedrigkontrastierte Läsionen in Phantome eingefügt werden, die anschließend für Bildbewertungen in Detektierbarkeitsexperimenten

verwendet werden können. In einer vorausgehenden Arbeit wurde in diesem Zusammenhang gezeigt, dass mit diesem Ansatz Halsphantome mit Niedrigkontrastläsionen für die detektionsbasierte Bewertung von Bildqualität hergestellt werden können und dass die Detektionsschwelle in anatomisch realistischen Halsphantomen zwischen 20 und 30 HE liegt (34). Damit lagen die Kontraste deutlich oberhalb der in homogenen Phantomen üblicherweise verwendeten Signalkontraste, die häufig im Bereich von 5 bis 15 HE liegen (35).

Diese Beobachtungen standen in guter Übereinstimmung mit einer vorausgehenden CT-Studie, in der komplexe Strukturen ebenfalls einen Einfluss auf die Detektionsraten von Niedrigkontrastläsionen durch mathematische Modelle besaßen (36). In dieser Studie bestanden darüber hinaus relevante Unterschiede im Einfluss von Dosis und Bildrekonstruktion zwischen einfachen homogenen und komplexen Hintergrundtexturen. Im Gegensatz hierzu wurden in einer jüngeren Studie, die ebenfalls die Detektion durch mathematische Modelle untersuchte, keine wesentlichen Unterschiede zwischen der Detektionsrate in einem homogenen Wasserhintergrund und einem als komplex eingestuften Leberhintergrund gefunden (29).

Vor diesem Hintergrund wurde im Rahmen der vorliegenden Arbeit (37) der Einfluss der Hintergrundtextur von Phantomen auf die Detektion von Niedrigkontrastläsionen und die Bewertung von Dosis und Bildrekonstruktion untersucht. Die Motivation der Arbeit ergab sich aus der Hypothese, dass die Hintergrundstruktur sowohl die Detektionsschwelle von Läsionskontrasten als auch Auswirkungen der Aufnahmedosis und des verwendeten Rekonstruktionsverfahrens auf den Detektionserfolg beeinflusst. Zur Untersuchung dieser Hypothese wurde ein Vergleich zwischen zwei formgleichen Phantomen mit homogener und anatomisch realistischer Hintergrundstruktur durchgeführt, die in vorausgehenden Arbeiten entwickelt und validiert wurden (34,38). Aufnahmen bei zwei verschiedenen Dosen und unter Verwendung von gefilterter Rückprojektion und iterativer Rekonstruktion wurden verwendet, um die Detektierbarkeit von Niedrigkontrastläsionen durch Radiologen und Radiologinnen zu bewerten. Darüber hinaus wurden Rausch- und Texturcharakteristika untersucht und zwischen den beiden Phantomarten verglichen (37).

## 1.2 Fragestellungen

Im Einzelnen wurden folgende Fragestellungen untersucht:

1. Bestehen signifikante Unterschiede in der Detektionsrate von Niedrigkontrastläsionen zwischen homogenen und anatomischen Phantomen bei vergleichbarem Läsionskontrast?
2. Bestehen in diesem Zusammenhang Unterschiede in der wahrgenommenen diagnostischen Sicherheit der teilnehmenden Radiologen und Radiologinnen?
3. Welchen Einfluss haben die Strahlendosis und Rekonstruktionsverfahren auf die Detektionsrate in homogenen Phantomen im Vergleich zu anatomischen Phantomen?
4. Bestehen zwischen beiden Phantomarten Unterschiede im Bildrauschen und in der Texturcharakterisierung in Abhängigkeit von Dosis und Bildrekonstruktion?



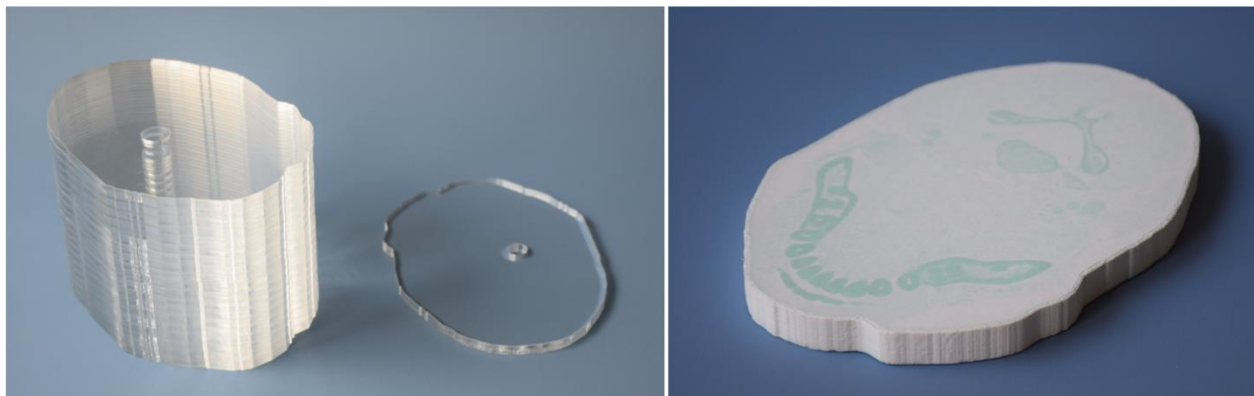
## 2 Methodik

### 2.1 In dieser Arbeit verwendete Phantome

In dieser Arbeit wurden formgleiche Halsphantome mit homogener und anatomischer Textur verwendet, die in vorausgehenden Arbeiten beschrieben und validiert wurden (34,38).

Beide Phantomvarianten wurden auf Grundlage desselben Hals-CT-Bildes einer Patientin hergestellt. Das homogene Phantom bestand aus Polymethylmethacrylat und wies eine zirkuläre Öffnung mit einem Durchmesser von 1 cm im linken parapharyngealen Raum auf. Durch die Platzierung eines mit Fruktoselösungen verschiedener Konzentrationen befüllten Angioplastieballons wurden Niedrigkontrastläsionen erzeugt. Das anatomische Phantom wurde mittels strahlendichtem 3D-Druck hergestellt (32) und die 1 cm messenden, zirkulären Niedrigkontrastläsionen wurden erzeugt, indem die Läsionen vor der Herstellung in das CT-Bild eingefügt wurden (34).

Insgesamt wurden fünf unterschiedliche Läsionskontraste im homogenen Phantom und vier Kontraste im anatomischen Phantom verwendet: 4, 9, 18, 30 und 38 HE (homogen) und 10, 18, 30 und 38 HE (anatomisch). Zudem wurde jeweils eine Phantomvariante ohne Läsion verwendet. Abbildung 1 zeigt Photographien der beiden Phantome.



**Abbildung 1:** Photographien der in dieser Arbeit verwendeten Phantome: Homogenes Phantom aus Polymethylmethacrylat (links) und Beispiel eines 3D-gedruckten anatomischen Phantoms (rechts). Quelle: Eigenes Bildmaterial.

## 2.2 CT-Aufnahmen

Die CT-Aufnahmen der homogenen Phantome lagen aus einer vorherigen Studie vor (38). Die Aufnahmen der anatomischen Phantome erfolgten mit einem Canon Aquilion Prime CT Scanner (Canon Medical Systems, Otawara, Japan) mit denselben Einstellungen. Die Röhrenspannung betrug 120 kVp und zwei Röhrenströme wurden verwendet: 30 und 120 mA. Der computed tomography dose index ( $CTDI_{vol}$ ) betrug 1,4 und 5,6 mGy. Die Aufnahmen erfolgten als Spiral-CT mit einer Rotationszeit von 0,5 s und einem Pitch von 0,813. Die Kollimation betrug 80 x 0,5 mm und das Sichtfeld betrug 28 cm im Durchmesser. Jede Aufnahme wurde fünfmal wiederholt. Die CT-Bilder wurden mit einer Schichtdicke von 0,5 mm und einem Weichteilkernel (FC08) rekonstruiert. Die Rekonstruktion erfolgte mit zwei Rekonstruktionsverfahren: Adaptive iterative Dosisreduktion 3D (AIDR 3D) und gefilterte Rückprojektion (FBP).

## 2.3 Auswahl der CT-Bilder

Aus jeder CT-Aufnahme der läsionshaltigen Phantome wurden je vier Bilder ausgewählt. Hierbei wurde darauf geachtet, dass die Bilder aus der Mitte des Phantoms stammten und keine Artefakte aufwiesen, die die Wiedererkennung durch die Studienteilnehmenden begünstigen und die Ergebnisse verzerren könnten.

Insgesamt wurden 640 läsionshaltige Bilder extrahiert:

2 Phantomarten x 4 Läsionskontraste x 2 Röhrenströme x 2 Rekonstruktionsmethoden x 5 Aufnahmewiederholungen x 4 ausgewählte Bilder = 640 Bilder

Zusätzlich wurden aus den jeweils 5 wiederholten Aufnahmen der läsionsfreien Phantome 12 Bilder nach den oben genannten Kriterien extrahiert.

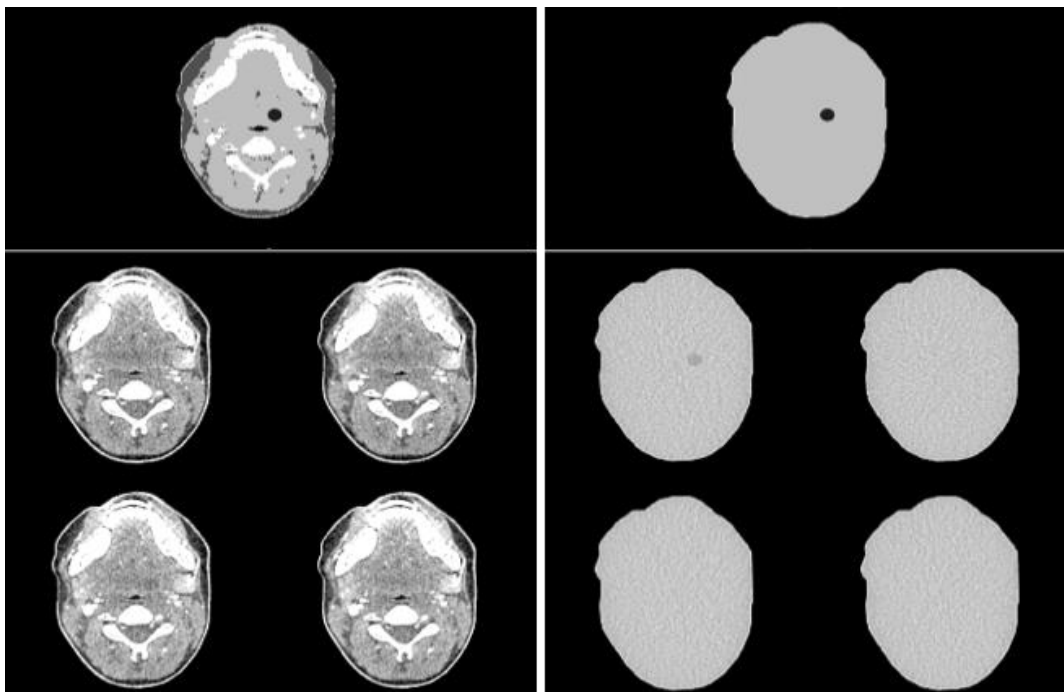
## 2.4 Aufbau des Detektionsexperimentes

Die Detektion von Niedrigkontrastläsionen wurde in einem 4-Alternative-Forced-Choice (4-AFC) -Experiment untersucht. Hierbei wurde den Teilnehmenden jeweils ein Quartett von vier CT-Bildern vorgelegt, von denen nur ein Bild eine Läsion enthielt. Die Studienteilnehmenden wurden gebeten, das Bild auszuwählen, das nach ihrer Auffassung vermutlich die Läsion enthielt. Alle vier Bilder eines Quartetts gehörten zur gleichen Phanto-

mart, wurden mit identischen CT-Einstellungen aufgenommen und mit derselben Methode rekonstruiert. Abbildung 2 illustriert den experimentellen Aufbau. Sowohl die Reihenfolge der Quartette als auch die Anzeigeposition des läsionshaltigen Bildes wurden randomisiert. Das 4-AFC-Experiment umfasste 320 Bildquartette mit Bildern des homogenen Phantoms und 320 Quartette mit Bildern des anatomischen Phantoms.

Sieben Radiologen und Radiologinnen nahmen an dem Experiment teil und bewerteten jeweils insgesamt 640 Bildquartette. Neben der Auswahl des läsionshaltigen Bildes wurden die Teilnehmenden gebeten, auch ihre Zuversicht bei der Entscheidung auf einer Skala von 1-5 anzugeben (1=sehr unsicher; 5=sehr sicher). Der Wert 1 war hierbei als minimale Sicherheit zu verstehen, sodass geraten werden musste. Der Wert 5 entsprach absoluter Sicherheit bei der Bildauswahl.

Um die Radiologen und Radiologinnen mit der Aufgabenstellung und der für die Auswertung verwendeten Softwareumgebung vertraut zu machen, führte jede teilnehmende Person vorab ein Trainingsexperiment mit je sieben Bildquartetten durch.



**Abbildung 2:** Aufbau des 4-AFC-Experiments. Beispiel für Bildquartette des anatomischen Phantoms (links) und des homogenen Phantoms (rechts). Über den Quartetten wird jeweils ein Referenzbild angezeigt, das die Position der zu detektierenden Läsion angibt. Quelle: Eigenes Bildmaterial.

## 2.5 Charakterisierung des Rauschens und Noise-Power-Spektrums

Das Rauschen und das Noise-Power-Spektrum (NPS) wurden in 200 Bildern pro Phantomtyp, Röhrenstromstärke und Rekonstruktionsmethode untersucht. Hierfür wurde in jedes Bild eine quadratische Maske (32x32 Pixel bzw. 17,5x17,5 mm) an derselben Stelle im Bild unmittelbar angrenzend an die Läsion gelegt. Es wurde darauf geachtet, dass die Maske im anatomischen Phantom in einer Region mit einer möglichst homogenen anatomischen Hintergrundstruktur lag. Gefäß- und Knochenstrukturen wurden ausgespart, um Verfälschungen durch besonders strahlendichte Strukturen zu vermeiden. Das Rauschen wurde als Standardabweichung der Pixelwerte berechnet. Das NPS wurde nach folgender Formel berechnet:

$$NPS(f_x, f_y) = \frac{b_x b_y}{L_x L_y} \langle |FFT_{2D}[ROI(x, y) - ROI_{Hintergrund}(x, y)]|^2 \rangle_{N_{ROI}}$$

wobei  $b_x$  und  $b_y$  die Pixelgröße (0.546 mm) und  $L_x$  und  $L_y$  die Kantenlänge der Masken waren (17.5 mm).  $FFT_{2D}$  war die 2D-Fourier-Transformation.  $ROI_{Hintergrund}$  bezeichnete die Maske zur Erfassung des Hintergrundrauschens.  $N_{ROI}$  war die Anzahl an verwendeten Masken (200) je Phantomart, Röhrenstrom und Rekonstruktionsmethode, aus denen die quadrierte Amplitude der Fourier-Transformation gemittelt wurde.

## 2.6 Datenanalyse

Die Detektionsrate ergab sich aus dem Verhältnis korrekt erkannter Läsionsbilder zu der Gesamtzahl der ausgewerteten Bildquartette je Phantomart, Läsionskontrast, Dosis, Bildrekonstruktion und Studienteilnehmenden. Mittels t-Test wurden die Detektionsraten und die Zuversicht bei der Entscheidung zwischen den Phantomarten verglichen. Der Vergleich von Dosis und Bildrekonstruktion erfolgte durch Varianzanalyse, wobei die p-Werte nach Tukey angepasst wurden. p-Werte unter 0,05 wurden als signifikant bewertet.

### 3 Ergebnisse

#### 3.1 Detektion von Niedrigkontrastläsionen im Vergleich zwischen homogenen und anatomischen Phantomen

In dem 4-AFC-Experiment zeigten sich bei allen Läsionskontrasten, Dosen und Rekonstruktionsmethoden signifikant niedrigere Detektionsraten und niedrigere diagnostische Zuversicht der Teilnehmenden in anatomischen als in homogenen Phantomen ( $p < 0,001$ ). Im homogenen Phantom wurde bereits bei niedrigen Kontrastwerten (9 HE) eine hohe Detektionsrate von  $89,5\% \pm 6,6\%$  (Mittelwert  $\pm$  Standardabweichung) erreicht. Bei 18 HE Läsionskontrast lag die Detektionsrate bereits bei  $99,6\% \pm 0,5\%$  und höhere Kontraste (30 und 38 HE) führten bei homogenem Phantomhintergrund zur erfolgreichen Detektion in allen Fällen. Im anatomischen Phantom hingegen waren die Detektionsraten bei 10 HE ( $52,9\% \pm 8,8\%$ ) und 18 HE ( $55,5\% \pm 8,3\%$ ) Läsionskontrast deutlich niedriger. Hohe bis annähernd perfekte Detektionsraten wurden erst ab einem Kontrast von 30 HE erreicht (30 HE:  $91,1\% \pm 5,3\%$ ; 38 HE:  $97,5\% \pm 1,7\%$ ). Ähnliche Detektionsraten ergaben sich im Vergleich zwischen den beiden Phantomarten bei 30 HE Läsionskontrast im anatomischen Phantom ( $91,1\% \pm 5,3\%$ ) und 9 HE Kontrast im homogenen Phantom ( $89,5\% \pm 6,6\%$ ).

### 3.2 Einfluss der Dosis und Rekonstruktionsmethode auf die Detektion von Niedrigkontrastläsionen

Die Tabellen 1 bis 4 geben eine detaillierte Übersicht der Detektionsraten und diagnostischen Zuversicht der Teilnehmenden für beide Phantome und die untersuchten Läsionskontraste, Dosen und Rekonstruktionsverfahren.

**Tabelle 1:** Detektionsraten (Prozent  $\pm$  Standardabweichung) im homogenen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: Modifiziert nach Conzelmann, J et al. 2021.

Homogenes Phantom				
CTDIvol	1,4 mGy		5,6 mGy	
Rekonstruktion	IR	FBP	IR	FBP
4 HE	67,9 $\pm$ 9,9	42,9 $\pm$ 15,6	70,7 $\pm$ 14	70 $\pm$ 14,1
9 HE	93,6 $\pm$ 9,2	67,9 $\pm$ 19,1	99,3 $\pm$ 1,8	97,1 $\pm$ 3,6
18 HE	100 $\pm$ 0	99,3 $\pm$ 1,8	100 $\pm$ 0	99,3 $\pm$ 1,8
30 HE	100 $\pm$ 0	100 $\pm$ 0	100 $\pm$ 0	100 $\pm$ 0
38 HE	100 $\pm$ 0	100 $\pm$ 0	100 $\pm$ 0	100 $\pm$ 0

**Tabelle 2:** Detektionsraten (Prozent  $\pm$  Standardabweichung) im anatomischen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: modifiziert nach Conzelmann, J et al. 2021.

Anatomisches Phantom				
CTDIvol	1,4 mGy		5,6 mGy	
Rekonstruktion	IR	FBP	IR	FBP
10 HE	47,9 $\pm$ 22,4	48,6 $\pm$ 22,3	66,4 $\pm$ 15	48,6 $\pm$ 22,3
18 HE	54,3 $\pm$ 17,8	45 $\pm$ 22	66,4 $\pm$ 23,2	56,4 $\pm$ 14,8
30 HE	90,7 $\pm$ 11,4	85,7 $\pm$ 19,5	92,1 $\pm$ 9,9	95,7 $\pm$ 6,7
38 HE	99,3 $\pm$ 1,8	93,6 $\pm$ 4,4	99,3 $\pm$ 1,8	97,9 $\pm$ 5,3

**Tabelle 3:** Diagnostische Zuversicht (Mittelwert  $\pm$  Standardabweichung) im homogenen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: Eigene Darstellung: Conzelmann, J.

	<b>Homogenes Phantom</b>			
<b>CTDIvol</b>	1,4 mGy		5,6 mGy	
<b>Rekonstruktion</b>	IR	FBP	IR	FBP
<b>4 HE</b>	1,59 $\pm$ 0,59	1,4 $\pm$ 0,53	1,6 $\pm$ 0,7	1,68 $\pm$ 0,77
<b>9 HE</b>	2,9 $\pm$ 1,2	1,9 $\pm$ 1	3,1 $\pm$ 1,1	2,9 $\pm$ 1
<b>18 HE</b>	4 $\pm$ 1	3,8 $\pm$ 1,1	4,4 $\pm$ 0,8	4,3 $\pm$ 0,9
<b>30 HE</b>	4,7 $\pm$ 0,4	4,5 $\pm$ 0,7	4,9 $\pm$ 0,2	4,9 $\pm$ 0,2
<b>38 HE</b>	4,9 $\pm$ 0,2	4,9 $\pm$ 0,2	5 $\pm$ 0,02	5 $\pm$ 0,1

**Tabelle 4:** Diagnostische Zuversicht (Mittelwert  $\pm$  Standardabweichung) im anatomischen Phantom je Läsionskontrast in Hounsfield Einheiten (HE), Rekonstruktionsmethode und computed tomography dose index (CTDIvol). IR = iterative Rekonstruktion, FBP = gefilterte Rückprojektion. Quelle: Eigene Darstellung: Conzelmann, J.

	<b>Anatomisches Phantom</b>			
<b>CTDIvol</b>	1,4 mGy		5,6 mGy	
<b>Rekonstruktion</b>	IR	FBP	IR	FBP
<b>10 HE</b>	1,3 $\pm$ 0,4	1,2 $\pm$ 0,4	1,6 $\pm$ 0,7	1,2 $\pm$ 0,4
<b>18 HE</b>	1,3 $\pm$ 0,4	1,3 $\pm$ 0,5	1,5 $\pm$ 0,6	1,3 $\pm$ 0,4
<b>30 HE</b>	2,2 $\pm$ 0,7	1,9 $\pm$ 0,8	2,3 $\pm$ 0,8	2,3 $\pm$ 1
<b>38 HE</b>	3,3 $\pm$ 1	2,8 $\pm$ 0,9	3,4 $\pm$ 0,9	3,1 $\pm$ 0,8

Zur Bewertung von Dosis- und Rekonstruktionseffekten wurden die Ergebnisse bei 4, 9 und 18 HE Läsionskontrast im homogenen Phantom und die Ergebnisse bei 10, 18, 30 und 38 HE Kontrast im anatomischen Phantom herangezogen. Die Ergebnisse im homogenen Kontext bei 30 und 38 HE Kontrast wurden nicht verwendet, da hier bereits in allen Fällen Detektionsraten von 100% vorlagen.

In den homogenen Phantomen zeigten sich signifikante Dosiseffekte bei Verwendung von FBP, die zu einem Abfall der Detektionsrate bei einer Dosisreduktion von 5,6 auf 1,4 mGy führten (5,6 mGy:  $83,6 \pm 10,2\%$  vs. 1,4 mGy:  $55,4\% \pm 12,9\%$ ;  $p < 0,001$ ). Bei iterativ rekonstruierten Bildern bestanden hingegen keine signifikanten Effekte der Dosis auf die Detektionsrate (5,6 mGy:  $85\% \pm 10,4\%$  vs. 1,4 mGy:  $80,7 \pm 9,6\%$ ;  $p = 0,375$ ). Im anatomischen Kontext waren die Unterschiede in den Detektionsraten zwischen 5,6 und 1,4 mGy weniger ausgeprägt, aber dennoch statistisch signifikant bei beiden Rekonstruktionsverfahren (FBP:  $74,6 \pm 22\%$  bei 5,6 mGy vs.  $68,2 \pm 21\%$  bei 1,4 mGy;  $p = 0,027$  und IR:  $81,1 \pm 0,14\%$  bei 5,6 mGy vs.  $73 \pm 22\%$  bei 1,4 mGy;  $p = 0,018$ ).

Der Vergleich der Rekonstruktionsverfahren zeigte eine klare Überlegenheit von IR in homogenen Phantomen bei 1,4 mGy ( $p < 0,001$ ), nicht jedoch bei 5,6 mGy ( $p = 0,785$ ). Bemerkenswerterweise bestätigten sich diese Ergebnisse im anatomischen Kontext nicht ( $p = 0,144$  bei 1,4 mGy und  $p = 0,111$  bei 5,6 mGy).

### 3.3 Rauscheigenschaften und Noise-Power-Spektrum

Niedrigdosisaufnahmen bei 1,4 mGy und FBP gingen in beiden Phantomarten mit den höchsten Rauschwerten einher (homogenes Phantom:  $26,3 \pm 1,3$ ; anatomisches Phantom:  $39,8 \pm 2,1$ ). Die niedrigsten Werte entstanden jeweils bei der Kombination aus 5,6 mGy Dosis und IR (homogenes Phantom:  $9,8 \pm 0,6$ ; anatomisches Phantom:  $23,5 \pm 1,7$ ). Das Rauschen war in CT-Bildern des anatomischen Phantoms unabhängig von der Rekonstruktionsmethode und Dosis höher als im homogenen Phantom ( $p < 0,001$ ).

Im homogenen Phantom führte eine Dosissteigerung von 1,4 auf 5,6 mGy unabhängig von der Rekonstruktionsmethode zu einer Reduktion des Rauschens ( $p < 0,001$ ). Bemerkenswerterweise traf dies im anatomischen Phantom nur für Bilder zu, die mit FBP rekonstruiert wurden ( $39,8 \pm 2,1$  bei 1,4 mGy vs.  $27,5 \pm 1,7$  bei 5,6 mGy;  $p < 0,001$ ), nicht jedoch für IR ( $23,8 \pm 1,6$  bei 1,4 mGy vs.  $23,5 \pm 1,7$  bei 5,6 mGy;  $p = 0,26$ ).

Bemerkenswert war auch, dass der Rauschanstieg durch die Dosisreduktion von 5,6 auf 1,4 mGy im anatomischen Phantom durch den Wechsel von FBP zu IR kompensiert wurde ( $27,5 \pm 1,7$  bei 5,6 mGy und FBP vs.  $23,8 \pm 1,6$  bei 1,4 mGy und IR;  $p < 0,001$ ). Diese Beobachtung wurde im homogenen Phantomhintergrund jedoch nicht bestätigt (Rauschwerte  $12,8 \pm 0,6$  bei 5,6 mGy und FBP vs.  $13,9 \pm 0,8$  bei 1,4 mGy und IR;  $p < 0,001$ ).



Bei 1,4 mGy und IR zeigte die NPS-Kurve im homogenen Phantomkontext eine Verschiebung der Maximalamplitude zu niedrigeren räumlichen Frequenzen mit dem NPS-Maximum bei einer Frequenz von  $0,23 \text{ mm}^{-1}$ . In den anatomischen Phantomen war der Frequenzwert des NPS-Maximums hingegen unabhängig von der Dosis und Rekonstruktionsmethode stets bei  $0,12 \text{ mm}^{-1}$ .

Tabellen 5 und 6 fassen die Ergebnisse der Rausch- und NPS-Auswertung für beide Phantomarten und die untersuchten Dosen und Rekonstruktionsmethoden zusammen.

**Tabelle 5:** Rauschwerte als Mittelwert  $\pm$  Standardabweichung der gemessenen Pixelstandardabweichungen. Quelle: Modifiziert nach Conzelmann, J et al. 2021

<b>CTDIvol &amp; Rekonstruktion</b>	<b>Homogenes Phantom</b>	<b>Anatomischen Phantom</b>
<b>1,4 mGy FBP</b>	$26,3 \pm 1,3$	$39,8 \pm 2,1$
<b>1,4 mGy AIDR 3D</b>	$13,9 \pm 0,8$	$23,8 \pm 1,6$
<b>5,6 mGy FBP</b>	$12,8 \pm 0,6$	$27,5 \pm 1,7$
<b>5,6 mGy AIDR 3D</b>	$9,8 \pm 0,6$	$23,5 \pm 1,7$

**Tabelle 6:** Räumliche Frequenz des Noise Power Spektrum Maximums ( $\text{mm}^{-1}$ ). Quelle: Modifiziert nach Conzelmann, J. et al 2021

<b>CTDIvol &amp; Rekonstruktion</b>	<b>Homogenes Phantom</b>	<b>Anatomisches Phantom</b>
<b>1,4 mGy FBP</b>	0,28	0,12
<b>1,4 mGy AIDR 3D</b>	0,23	0,12
<b>5,6 mGy FBP</b>	0,29	0,12
<b>5,6 mGy AIDR 3D</b>	0,28	0,12

## 4 Diskussion

### 4.1 Allgemeine Zielstellung

In der CT-Diagnostik besteht das Ziel nach dem „ALARA“-Prinzip darin, einen optimalen Kompromiss zwischen der notwendigen Strahlendosis und einer für die Diagnostik ausreichenden Bildqualität zu erzielen. Die Bewertung der Bildqualität kann anhand von objektiven physikalischen Parametern wie dem Bildrauschen, Kontrast oder Kontrast-zu-Rausch-Verhältnis erfolgen. Allerdings sind diese Parameter insbesondere im Zusammenhang mit nicht-linearen Rekonstruktionsverfahren wie iterativer Rekonstruktion nicht geeignet, um schwerwiegende Einschränkungen der diagnostischen Bildleistung auszuschließen. Hierfür können Methoden verwendet werden, die diagnostische Prozesse aus der klinischen Routine nachahmen (7). Ein häufiger Ansatz besteht dabei darin, die Detektionsrate von Niedrigkontrastläsionen zu bewerten. Zu diesem Zweck werden Phantome benötigt, die einen Vergleich der Detektionsergebnisse mit der Grundwahrheit, das heißt dem realen Vorhandensein von Läsionen, ermöglichen.

Phantome ermöglichen darüber hinaus im Gegensatz zu Patienten und Patientinnen wiederholte Aufnahmen ohne Bedenken hinsichtlich der Risiken von ionisierender Strahlung (4). Dies ist beispielsweise vorteilhaft, um den Einfluss von verschiedenen CT-Einstellungen und Rekonstruktionsmethoden auf die Detektionsrate gegenüberstellen zu können. Bislang kommen vorwiegend homogene Phantome zum Einsatz und es wird im Allgemeinen angenommen, dass sich die Ergebnisse aus der Bewertung von CT-Technologien auf die klinische Diagnostik von Patienten und Patientinnen übertragen lassen (24). Allerdings bestehen aus vorausgehenden Arbeiten deutliche Hinweise darauf, dass dies nicht zutrifft und Phantomeigenschaften die Bewertung und Schlussfolgerungen im Hinblick auf zielgerechte Bildgebung in der klinischen Routine beeinflussen. Vor diesem Hintergrund bestand das Ziel dieser Arbeit darin, einen direkten Vergleich zwischen homogenen und realistischen, anatomischen Phantomen durchzuführen und den Einfluss der Phantomstruktur auf die Detektierbarkeit von Niedrigkontrastläsionen und die Bewertung von CT-Technologien zu untersuchen. Zu diesem Zweck wurden formgleiche Halsphantome mit homogener und anatomischer Textur und Niedrigkontrastläsionen mit 4 bis 38 HE Kontrast bei zwei Dosen untersucht. Die Bilder wurden mit gefilterter Rückprojektion und iterativer Rekonstruktion rekonstruiert und die Aufnahmen wurden in einem Detekti-

onsexperiment durch sieben Radiologen und Radiologinnen ausgewertet. Darüber hinaus wurden Rausch- und Textureigenschaften zwischen den beiden Phantomtypen verglichen (37).

#### **4.2 Zusammenfassung der wichtigsten Ergebnisse und Einordnung in die Literatur**

In vorausgehenden Arbeiten wurde berichtet, dass die Komplexität von Phantomen und CT-Bildern die Detektionsrate von Niedrigkontrastläsionen beeinflusst (25,39). Diese Einschätzung konnte durch unsere Ergebnisse bestätigt werden (37). Im Vergleich zwischen anatomisch realistischen und homogenen Halsphantomen zeigte sich eine signifikant geringere Detektionsrate von Läsionen im anatomischen Kontext. Dieser deutliche Unterschied traf für alle Läsionskontraste und bei allen untersuchten Dosen und Rekonstruktionsverfahren zu. Um vergleichbare Detektionsraten zwischen den beiden Phantomvarianten zu erzielen, waren deutlich höhere Läsionskontraste im anatomischen Hintergrund notwendig (30 HE vs. 9 HE im homogenen Phantom). Im homogenen Phantom wurde von den Studienteilnehmenden bereits bei niedrigen Läsionskontrasten von 18 HE fast perfekte Detektionsraten erzielt, die im anatomischen Phantom erst bei 38 HE erreicht wurden (37).

Die von uns beobachteten, relativ hohen Detektionsraten in homogenen Phantomen waren in guter Übereinstimmung mit den Ergebnissen vorausgehender Studien (16,40). Weiterhin bestätigten unsere Ergebnisse (37) im homogenen Kontext auch das vorausgehend anhand von ähnlichen homogenen Phantomen beschriebene Potenzial zur Dosisreduktion durch iterative Rekonstruktion (20,41,42). Diese Schlussfolgerung wurde in den anatomischen Phantomen jedoch nicht bestätigt. Hier zeigte sich kein Vorteil von iterativer Rekonstruktion gegenüber gefilterter Rückprojektion bei niedrigen Dosen. Unsere Ergebnisse sind damit in guter Übereinstimmung mit einer Studie von Solomon et al., in der komplexe Hintergrundstrukturen ebenfalls die Bewertung von iterativer Rekonstruktion beeinflussten (36).

Darüber hinaus hatte die Hintergrundstruktur der untersuchten Phantome auch Einflüsse auf die Bewertung des Bildrauschens und der Rauschtextur. So war das Rauschen im anatomischen Phantom konsistent höher als im homogenen Phantom. Bemerkenswert war darüber hinaus, dass das Rauschen bei iterativer Rekonstruktion in homogenen Bildern mit sinkender Dosis anstieg und dass dieser Effekt in anatomischen Bildern nicht zu

beobachten war. Dies deutet darauf hin, dass iterative Rekonstruktion im anatomischen Kontext bereits bei niedrigen Dosen ihre Leistung ausspielen kann, um effektiv minimales Bildrauschen zu erzielen. Im Gegensatz zu den homogenen Phantomen wurde in den anatomischen Phantomen ein Anstieg der NPS-Kurve bei niedrigen Frequenzen registriert, der vorausgehenden Ergebnissen aus CT-Aufnahmen von Patienten und Patientinnen ähnelte (43). Diese Beobachtung illustriert die Relevanz anatomischer Textur und deutet auf eine adäquate Vergleichbarkeit der hier verwendeten Phantome mit klinischen Aufnahmen hin.

Über relevante Einflüsse von anatomischen Strukturen und realistischen Texturen wurde bereits in vorausgehenden Studien berichtet (25,26,44). So wurde in einer Arbeit beispielsweise die Detektionsrate von sehr kleinen Läsionen stärker vom anatomischen Rauschen als von dem Quantenrauschen beeinflusst (25). In einer weiteren Studie ging eine Dosissteigerung zur Reduktion des Bildrauschens aufgrund des Einflusses anatomischer Strukturen nicht mit einer verbesserten Detektionsrate einher (44). Es ist davon auszugehen, dass die Detektion von Läsionen im anatomischen Kontext generell limitierter als in vereinfachten, homogenen Bildern ist (27). Insgesamt passen unsere Ergebnisse (37) zu der vorbeschriebenen Komplexität und den zugrunde liegenden Interaktionen zwischen anatomischer Struktur, Rauschen und räumlicher Auflösung (25–27,39,44).

Im Gegensatz dazu stehen die Ergebnisse einer weiteren Studie, in der sich kein Einfluss der Hintergrundtextur auf die Detektionsrate von Niedrigkontrastläsionen zeigte (29). Dieser Dissens lässt sich jedoch durch Unterschiede im Versuchsaufbau erklären. In der vorausgehenden Studie von Dilger et al. wurden CT-Bilder eines homogenen Wasserphantoms mit Aufnahmen von gefäßfreien Bildausschnitten von Leberparenchym verglichen (29). Dabei ähnelten sich die gegenübergestellten Hintergrundstrukturen jedoch stark, wodurch sich möglicherweise die fehlenden Effekte der Hintergrundstruktur erklären lassen. Im Gegensatz hierzu bildete das in der vorliegenden Arbeit verwendete Phantom die Halsanatomie detailgetreu nach und wies eine deutlich höhere Komplexität als das gefäßfreie Leberparenchym auf. Unsere Arbeit unterschied sich auch grundlegend durch den Herstellungsprozess der CT-Bilder (37). Dilger et al. fügten Läsionen nachträglich digital in CT-Aufnahmen von Leberparenchym ein und simulierten CT-Einstellungen durch digitale Bildbearbeitung. Im Gegensatz hierzu lagen in der vorliegenden Arbeit reale Aufnahmen von Phantomen mit realen Niedrigkontrastläsionen vor, wodurch sich

Fehlerquellen in der digitalen Manipulation und Verarbeitung CT-Bildern vermeiden ließen.

### **4.3 Limitationen der Studie**

Die vorliegende Arbeit zur Untersuchung des Einflusses der Hintergrundtextur auf die Detektion von Niedrigkontrastläsionen hatte mehrere Limitationen (37). Hierzu gehört, dass nur eine einzige Läsionsgröße und eine Läsionsform untersucht wurden. Die Läsionen befanden sich immer an der gleichen Stelle und es wurde ein experimentelles Vorgehen gewählt, bei dem die Studienteilnehmenden Kenntnis über die zu erwartende Position besaßen. Allerdings wurde dieses Studiendesign bewusst gewählt, da unterschiedliche Positionen potenziell die Vergleichbarkeit der Ergebnisse ungünstig beeinflusst hätten. Des Weiteren war die Anzahl der Läsionskontraste limitiert. Sehr kontrastarme Läsionen wie z. B. 5 HE wurden in anatomischen Phantomen nicht untersucht, da sich wie aus vorausgehenden Arbeiten bekannt bereits bei 10 HE Kontrast niedrige Detektionsraten ergeben (34). Darüber hinaus war auch der Umfang der untersuchten Dosiseinstellungen und Rekonstruktionsverfahren begrenzt und es wurde nur ein einziges CT-Gerät für die Aufnahmen verwendet. Es besteht die Möglichkeit, dass die Ergebnisse bei anderen Akquisitionseinstellungen und CT-Geräten abweichen. Das Ziel der Arbeit bestand jedoch nicht darin, umfassend Aufnahmetechniken zu untersuchen, sondern Einflüsse der Hintergrundstruktur von Phantomen anhand eines umschriebenen Versuchsaufbaus und unter vergleichbaren Bedingungen zu untersuchen. Weiterführende Untersuchungen beispielsweise zur Optimierung von Aufnahmeprotokollen sind in anderen Studien zu finden (27,45).

### **4.4 Implikationen für Praxis und/oder zukünftige Forschung**

Nach aktuellem Wissen wurden im Rahmen der vorliegenden Arbeit erstmals Unterschiede in der Detektion von Niedrigkontrastläsionen und darauf aufbauenden Bewertung von CT-Technologien zwischen homogenen und anatomisch realistischen Phantomen untersucht. Hierbei wurden signifikante Einflüsse der Hintergrundtextur beobachtet. Hieraus ergeben sich weitreichende Implikationen für die Bewertung von CT-Technologien und die Übertragbarkeit von Ergebnissen aus dem experimentellen Umfeld auf die klinische Bildgebung von Patienten und Patientinnen. Als allgemeine Schlussfolgerung kann abgeleitet werden, dass Phantome der Komplexität des menschlichen Körpers umso

mehr gerecht werden, je realistischer sie die Anatomie und Gewebedetails des Menschen widerspiegeln. Somit hat die vorliegende Arbeit einen grundlegenden Beitrag zur Optimierung von Bildgebung gemäß dem ALARA-Prinzip geleistet.

## **5 Schlussfolgerungen**

Phantome spielen eine zentrale Rolle in der Entwicklung und Bewertung von CT-Technologien. In der vorliegenden Arbeit wurden homogene und anatomische Phantomvarianten verglichen und im Hinblick auf die Detektionsrate von niedrig kontrastierten Läsionen und Rausch- und Textureigenschaften untersucht. Hierbei zeigten sich signifikante Unterschiede im Detektionserfolg bei vergleichbaren Kontrasten und es zeigte sich, dass die Hintergrundtextur die Bewertung von Dosis und Rekonstruktionsverfahren beeinflusst. Hieraus ergeben sich weitreichende Konsequenzen für die Bewertung von CT-Bildgebung in Phantom-Experimenten, die klinische Bildgebung möglichst realistisch wiedergeben sollten, um Ergebnisse zu generieren, die auf die Diagnostik von Patienten und Patientinnen übertragbar sind.

## Literaturverzeichnis

1. Bundesamt für Strahlenschutz. Forschen - Informieren - Schützen. 2017 [cited 2022 Jan 24]; Available from: [www.bfs.de](http://www.bfs.de)
2. Hounsfield GN. Computerized transverse axial scanning (tomography): I. Description of system. *Br J Radiol.* 1973;46(552):1016-22.
3. Prokop M. Strahlendosis in der Computertomographie. Risiko und Herausforderung. Vol. 48, *Radiologe*. Springer; 2008. p. 229–42.
4. Shah DJ, Sachs RK, Wilson DJ. Radiation-induced cancer: A modern view. *British Journal of Radiology.* 2012;85(1020):e1166-73.
5. Brenner DJ. Estimating cancer risks from pediatric CT: Going from the qualitative to the quantitative. *Pediatr Radiol.* 2002;32(4):228-1; discussion 242-4.
6. Mountford PJ, Temperton DH. Recommendations of the International Commission on Radiological Protection (ICRP) 1990. *European Journal of Nuclear Medicine.* 1992;19(2):77-9.
7. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Med Phys.* 2014; 41(7):071904.
8. Baker ME, Dong F, Primak A, Obuchowski NA, Einstein D, Gandhi N, Herts BR, Purysko A, Remer E, Vachhani N. 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. *Am J Roentgenol.* 2012;199(1):8–18.
9. Hara AK, Paden RG, Silva AC, Kujak JL, Lawder HJ, Pavlicek W. Iterative reconstruction technique for reducing body radiation dose at CT: Feasibility study. *Am J Roentgenol.* 2009;193(3):764–71.

10. Tang K, Wang L, Li R, Lin J, Zheng X, Cao G. Effect of low tube voltage on image quality, radiation dose, and low-contrast detectability at abdominal multidetector CT: Phantom study. *J Biomed Biotechnol.* 2012;2012:130169.
11. Alkadhi H, Leschka S, Stolzmann P, Scheffel H. *Wie funktioniert CT? Protokollparameter und Bildqualität.* Springer Verlag. 2011;
12. Bushberg JT, Seibert JA, Leidholdt EM, Boone JM. *The essential physics of medical imaging.* 1048 p.
13. Jones J, Murphy A. Radiographic contrast. *Radiopaedia.org* [Internet]. 2018 Mar 3 [cited 2022 Jan 24]; Available from: <http://radiopaedia.org/articles/58718>
14. Hounsfield unit | Radiology Reference Article | *Radiopaedia.org* [Internet]. [cited 2022 Jun 28]. Available from: <https://radiopaedia.org/articles/hounsfield-unit>
15. Mileto A, Guimaraes LS, McCollough CH, Fletcher JG, Yu L. State of the art in abdominal CT: The limits of iterative reconstruction algorithms. *Radiology.* 2019;293(3):491-503.
16. Euler A, Stieltjes B, Szucs-Farkas Z, Eichenberger R, Reisinger C, Hirschmann A, Zaehringer C, Kircher A, Streif M, Bucher S, Buergler D, D'Errico L, Kopp S, Wilhelm M, Schindera ST. Impact of model-based iterative reconstruction on low-contrast lesion detection and image quality in abdominal CT: a 12-reader-based comparative phantom study with filtered back projection at different tube voltages. *Eur Radiol.* 2017;27(12):5252–9.
17. Cormack AM. Reconstruction of densities from their projections, with applications in radiological physics. *Phys Med Biol.* 1973;18(2):195-207.
18. Solomon J, Marin D, Choudhury KR, Patel B, Samei E. Effect of radiation dose reduction and reconstruction algorithm on image noise, contrast, resolution, and detectability of subtle hypoattenuating liver lesions at Multidetector CT: Filtered back projection versus a commercial model-based iterative reconstruction algorithm. *Radiology.* 2017; 284(3):777–87.



19. Schindera ST, Diedrichsen L, Müller HC, Rusch O, Marin D, Schmidt B, Raupach R, Vock P, Szucs-Farkas Z. Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: Assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study. *Radiology*. 2011;260(2):454–62.
20. Xu Y, Zhang TT, Hu ZH, Li J, Hou HJ, Xu ZS, He W. Effect of iterative reconstruction techniques on image quality in low radiation dose chest CT: A phantom study. *Diagnostic Interv Radiol*. 2019;25(6):442-450.
21. Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. *Physica Medica*. 2012;28(2):94-108.
22. Bellesi L, Wyttenbach R, Gaudino D, Colleoni P, Pupillo F, Carrara M, Braghetti A, Puligheddu C, Presilla S. A simple method for low-contrast detectability, image quality and dose optimisation with CT iterative reconstruction algorithms and model observers. *Eur Radiol Exp*. 2017;1(1):18.
23. Dolly S, Chen HC, Anastasio M, Mutic S, Li H. Practical considerations for noise power spectra estimation for clinical CT scanners. *J Appl Clin Med Phys*. 2016;17(3):392-407.
24. Husby E, Svendsen ED, Andersen HK, Martinsen ACT. 100 days with scans of the same Catphan phantom on the same CT scanner. *J Appl Clin Med Phys*. 2017; 18(6):224-231.
25. Kotre CJ. The effect of background structure on the detection of low contrast objects in mammography. *Br J Radiol*. 1998;71(851):1162-7.
26. Samei E, Flynn MJ, Eyler WR. Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs. *Radiology*. 1999;213(3):727-34.
27. Smith TB, Rubin GD, Solomon J, Harrawood B, Choudhury KR, Samei E. Local complexity metrics to quantify the effect of anatomical noise on detectability of lung nodules in chest CT imaging. *J Med Imaging*. 2018;5(4):045502.

28. De Crop A, Smeets P, Van Hoof T, Vergauwen M, Dewaele T, Van Borsel M, Achten E, Verstraete K, D'Herde K, Thierens H, Bacher K. Correlation of clinical and physical-technical image quality in chest CT: A human cadaver study applied on iterative reconstruction. *BMC Med Imaging*. 2015;15:32.
29. Dilger SKN, Yu L, Chen B, Favazza CP, Carter RE, Fletcher JG, McCollough CH, Leng S. Localization of liver lesions in abdominal CT imaging: I. Correlation of human observer performance between anatomical and uniform backgrounds. *Phys Med Biol*. 2019;64(10):105011.
30. Einstein SA, Rong XJ, Jensen CT, Liu X. Quantification and homogenization of image noise between two CT scanner models. *J Appl Clin Med Phys*. 2019;21(1):174–178.
31. Gavrielides MA, Zeng R, Kinnard LM, Myers KJ, Petrick N. A template-based approach for the analysis of lung nodules in a volumetric CT phantom study. In: *Medical Imaging 2009: Computer-Aided Diagnosis*. 2009. p. 726009.
32. Jahnke P, Limberg FRP, Gerbl A, Pardo GLA, Braun VPB, Hamm B, Scheel M. Radiopaque three-dimensional printing: A method to create realistic CT phantoms. *Radiology*. 2017;282(2):569–75.
33. Jahnke P, Schwarz S, Ziegert M, Schwarz FB, Hamm B, Scheel M. Paper-based 3D printing of anthropomorphic CT phantoms: Feasibility of two construction techniques. *Eur Radiol*. 2019;29(3):1384-1390.
34. Ardila Pardo GL, Conzelmann J, Genske U, Hamm B, Scheel M, Jahnke P. 3D printing of anatomically realistic phantoms with detection tasks to assess the diagnostic performance of CT images. *Eur Radiol*. 2020;30(8):4557-4563.
35. 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. CT Detectability of Small Low-Contrast Hypoattenuating Focal Lesions: Iterative Reconstructions versus Filtered Back Projection. *Radiology*. 2018; 289(2):443-454.

36. Solomon J, Ba A, Bochud F, Samei E. Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms. *Med Phys*. 2016;43(12):6497.
37. Conzelmann J, Genske U, Emig A, Scheel M, Hamm B, Jahnke P. Comparison of low-contrast detectability between uniform and anatomically realistic phantoms-influences on CT image quality assessment. *Eur Radiol*. 2021;32(2):1267-1275.
38. Conzelmann J, Schwarz FB, Hamm B, Scheel M, Jahnke P. Development of a method to create uniform phantoms for task-based assessment of CT image quality. *J Appl Clin Med Phys*. 2020;21(9):201-208.
39. Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. *Med Phys*. 2014;41(9):091908.
40. Schindera ST, Odedra D, Raza SA, Kim TK, Jang HJ, Szucs-Farkas Z, Rogalla P. Iterative reconstruction algorithm for CT: Can radiation dose be decreased while low-contrast detectability is preserved? *Radiology*. 2013;269(2):511–8.
41. Raslau FD, Escott EJ, Smiley J, Adams C, Feigal D, Ganesh H, Wang C, Zhang J. Dose reduction while preserving diagnostic quality in head CT: Advancing the application of iterative reconstruction using a live animal model. *Am J Neuroradiol* 2019;40(11):1864–1870.
42. Lim K, Kwon H, Cho J, Oh J, Yoon S, Kang M, Ha D, Lee J, Kang E. Initial phantom study comparing image quality in computed tomography using adaptive statistical iterative reconstruction and new adaptive statistical iterative reconstruction v. *J Comput Assist Tomogr*. 2015;39(3):443–8.
43. Ba A, Abbey CK, Racine D, Viry A, Verdun FR, Schmidt S, Bochud FO. Channelized Hotelling observer correlation with human observers for low-contrast detection in liver CT images. *J Med Imaging*. 2019;6(2):025501.

44. Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P. Estimation of the noisy component of anatomical backgrounds. *Med Phys*. 1999;26(7):1365-70.
45. Jahnke P, Conzelmann J, Genske U, Nunninger M, Scheel M, Hamm B, Diekhoff T. Task-based assessment of neck CT protocols using patient-mimicking phantoms—effects of protocol parameters on dose and diagnostic performance. *Eur Radiol*. 2021;31(5):3177-3186.

## Eidesstattliche Versicherung

„Ich, Juliane Conzelmann, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Vergleich zwischen homogenen und realistischen Phantomen zur Bewertung von Bildqualität in der Computertomographie – „Vergleich zwischen homogenen und realistischen Phantomen zur Bewertung von Bildqualität in der Computertomographie-Phantomeinfluss auf Niedrigkontrastdetektierbarkeit / Comparison of uniform and realistic phantoms for image quality - evaluation in computed tomography – phantom effects on low-contrast detectability“

selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; [www.icmje.org](http://www.icmje.org)) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

## Anteilserklärung an den erfolgten Publikationen

Juliane Conzelmann hatte folgenden Anteil an der folgenden Publikation:

### Publikation 1:

Comparison of low-contrast detectability between uniform and anatomically realistic phantoms-influences on CT image quality assessment

Juliane Conzelmann, Ulrich Genske, Arthur Emig, Michael Scheel, Bernd Hamm, Paul Jahnke

European Radiology September 2021

### Beitrag im Einzelnen:

1. Durchführung der CT-Aufnahmen der homogenen und anatomischen Phantome in Zusammenarbeit mit PD Dr. Paul Jahnke
2. Management der Bilder
  - a) Sortierung und Auswahl der CT-Bilder
  - b) Nachbearbeitung der Bilder
    - a. Entfernung der Liege und anderer umliegender Artefakte
    - b. Zentrierung der Bilder
3. Erstellung des 4-AFC-Experiments und Trainingsstudien
  - a) Randomisierung der Bildreihenfolge
  - b) Erstellung von Trainingsstudien mit der Software HON
  - c) Erstellung der Hauptstudie mit der Software HON
4. Betreuung der 4-AFC-Experimente
  - a) Akquirierung und Betreuung der Teilnehmenden
  - b) Durchführung der Trainingsstudie mit 7 Radiologen und Radiologinnen
  - c) Durchführung der 4-AFC-Experimente mit 7 Radiologen und Radiologinnen
5. Statistische Auswertung der beiden 4-AFC-Experimente
  - a) Export der Ergebnisse zur Datenverarbeitung
  - b) Datenaufarbeitung für die statistische Auswertung
  - c) Vergleich der Detektionsrate und subjektiven Diagnosesicherheit in homogenen und anatomischen Phantomen

- d) Aus der statistischen Auswertung sind die Tabellen 1-6 entstanden
- 6. Erhebung der Rauschwerte in den Bildern
- 7. Literaturrecherche
  - a) Eigenständige Einarbeitung in den Forschungsstand zur Bewertung von CT-Bildqualität
- 8. Beteiligung an der Ausarbeitung und Revision des Manuskripts mit PD Dr. Paul Jahnke
  - a) Eigenständige Ausarbeitung von Abbildung 2
  - b) Unterstützung bei der Ausarbeitung von Abbildungen 4, 5 und 6
  - c) Manuskriptrevision
- 9. Veröffentlichung des Artikels in *European Radiology* September 2021

---

Unterschrift, Datum und Stempel des erstbetreuenden Hochschullehrers

---

Unterschrift der Doktorandin

## Auszug aus der Journal Summary List

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
21	EUROPEAN RADIOLOGY	20,761	4.101	0.033260
22	SEMINARS IN RADIATION ONCOLOGY	2,531	4.076	0.003540
23	JOURNAL OF MAGNETIC RESONANCE IMAGING	17,046	3.954	0.024900
24	Biomedical Optics Express	11,090	3.921	0.025030
25	COMPUTERIZED MEDICAL IMAGING AND GRAPHICS	2,656	3.750	0.002940
26	JOURNAL OF DIGITAL IMAGING	2,494	3.697	0.003790
27	MAGNETIC RESONANCE IN MEDICINE	32,159	3.635	0.029700
28	Insights Into Imaging	1,948	3.579	0.003260
29	INTERNATIONAL JOURNAL OF HYPERTHERMIA	4,397	3.574	0.004880
30	SEMINARS IN NUCLEAR MEDICINE	2,194	3.544	0.002420
31	AMERICAN JOURNAL OF NEURORADIOLOGY	23,135	3.381	0.027120
32	JOURNAL OF NUCLEAR CARDIOLOGY	3,600	3.366	0.004570
33	MEDICAL PHYSICS	26,445	3.317	0.027280
34	Quantitative Imaging in Medicine and Surgery	1,335	3.226	0.002800
35	NMR IN BIOMEDICINE	7,537	3.221	0.011610
36	Clinical Neuroradiology	935	3.183	0.002710
37	KOREAN JOURNAL OF RADIOLOGY	2,967	3.179	0.004490
38	Ultrasonography	618	3.075	0.001710
39	ULTRASONICS	7,808	3.065	0.008930
40	JOURNAL OF VASCULAR AND INTERVENTIONAL RADIOLOGY	9,045	3.037	0.009790
41	AMERICAN JOURNAL OF ROENTGENOLOGY	32,209	3.013	0.024770
42	Practical Radiation Oncology	1,879	2.948	0.005780



# Druckexemplar(e) der Publikation(en)

European Radiology (2022) 32:1267–1275  
<https://doi.org/10.1007/s00330-021-08248-3>

COMPUTED TOMOGRAPHY



## Comparison of low-contrast detectability between uniform and anatomically realistic phantoms—influences on CT image quality assessment

Juliane Conzelmann<sup>1</sup> · Ulrich Genske<sup>1,2</sup> · Arthur Emig<sup>1</sup> · Michael Scheel<sup>3</sup> · Bernd Hamm<sup>1</sup> · Paul Jahnke<sup>1,4</sup>

Received: 19 May 2021 / Revised: 22 July 2021 / Accepted: 5 August 2021 / Published online: 2 September 2021  
 © The Author(s) 2021

### Abstract

**Objectives** To evaluate the effects of anatomical phantom structure on task-based image quality assessment compared with a uniform phantom background.

**Methods** Two neck phantom types of identical shape were investigated: a uniform type containing 10-mm lesions with 4, 9, 18, 30, and 38 HU contrast to the surrounding area and an anatomically realistic type containing lesions of the same size and location with 10, 18, 30, and 38 HU contrast. Phantom images were acquired at two dose levels (CTDIvol of 1.4 and 5.6 mGy) and reconstructed using filtered back projection (FBP) and adaptive iterative dose reduction 3D (AIDR 3D). Detection accuracy was evaluated by seven radiologists in a 4-alternative forced choice experiment.

**Results** Anatomical phantom structure impaired lesion detection at all lesion contrasts ( $p < 0.01$ ). Detectability in the anatomical phantom at 30 HU contrast was similar to 9 HU contrast in uniform images (91.1% vs. 89.5%). Detection accuracy decreased from 83.6% at 5.6 mGy to 55.4% at 1.4 mGy in uniform FBP images ( $p < 0.001$ ), whereas AIDR 3D preserved detectability at 1.4 mGy (80.7% vs. 85% at 5.6 mGy,  $p = 0.375$ ) and was superior to FBP ( $p < 0.001$ ). In the assessment of anatomical images, superiority of AIDR 3D was not confirmed and dose reduction moderately affected detectability (74.6% vs. 68.2%,  $p = 0.027$  for FBP and 81.1% vs. 73%,  $p = 0.018$  for AIDR 3D).

**Conclusions** A lesion contrast increase from 9 to 30 HU is necessary for similar detectability in anatomical and uniform neck phantom images. Anatomical phantom structure influences task-based assessment of iterative reconstruction and dose effects.

### Key Points

- A lesion contrast increase from 9 to 30 HU is necessary for similar low-contrast detectability in anatomical and uniform neck phantom images.
- Phantom background structure influences task-based assessment of iterative reconstruction and dose effects.
- Transferability of CT assessment to clinical imaging can be expected to improve as the realism of the test environment increases.

**Keywords** Tomography, X-ray computed · Phantoms, imaging · Medical physics · Neck · Radiation protection

### Abbreviations

4-AFC 4-Alternative forced choice  
 AIDR 3D Adaptive iterative dose reduction 3D

CT Computed tomography  
 CTDI Computed tomography dose index

✉ Paul Jahnke  
[paul.jahnke@charite.de](mailto:paul.jahnke@charite.de)

<sup>1</sup> Department of Radiology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany

<sup>2</sup> Data Analytics and Computational Statistics, Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, 14482 Potsdam, Germany

<sup>3</sup> Department of Neuroradiology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany

<sup>4</sup> Berlin Institute of Health (BIH), Anna-Louisa-Karsch-Str. 2, 10178 Berlin, Germany

FBP Filtered back projection  
 HU Hounsfield unit

quality assessment in comparison with a uniform phantom background.

## Introduction

Image properties of clinical computed tomography (CT) images vary significantly due to differences between vendors, scanner generations, software versions, imaging techniques, and reconstruction methods. This diversity affects the diagnostic quality of CT images [1], and differences are likely to increase further as CT techniques evolve. In light of this situation, it is of relevance to ensure objective assessment and comparison of the clinical performance of CT techniques [2]. Task-based methods have been proposed for that purpose and should be applicable to evaluate the diagnostic performance of CT images regardless of the underlying imaging technology used [3, 4].

Task-based assessment is typically used to test lesion detectability in CT images of uniform phantoms, and it is commonly assumed that the results can be transferred to CT images of patients acquired in the clinical setting. Yet, there is evidence that uniform phantoms may not reflect clinical performance adequately. First, previous X-ray studies have shown that background structure affects detectability and conclusions about dose effects on image quality [5–7]. Second, background texture has also been identified to affect detectability and estimated dose reduction potential of an iterative reconstruction algorithm in a CT study [8]. Conversely, the authors of another CT study report only negligible texture effects, concluding that uniform phantoms may allow sufficient assessment of clinical performance [9]. Both of these CT studies investigated cropped images mimicking vessel-free liver textures. In order to better understand the validity of CT assessment with uniform phantoms for clinical imaging, it would be desirable to evaluate how such assessments relate to CT images obtained in phantoms with full anatomical detail.

A recent study introduced anatomically realistic neck phantoms that can be used for such purposes [10]. The phantoms investigated in that study contained low-contrast lesions and were produced using radiopaque 3D printing based on a neck CT image of a patient. Another recent study used the same CT image as a template to produce a uniform neck phantom for low-contrast detectability experiments [11]. The present study compares low-contrast detectability between these two types of phantoms to test the hypothesis that anatomical detail affects task-based CT assessment. CT images of the phantoms acquired at two dose levels and reconstructed with filtered back projection and an iterative reconstruction algorithm were analyzed. The overall aim was to evaluate the effects of anatomical background structure on task-based image

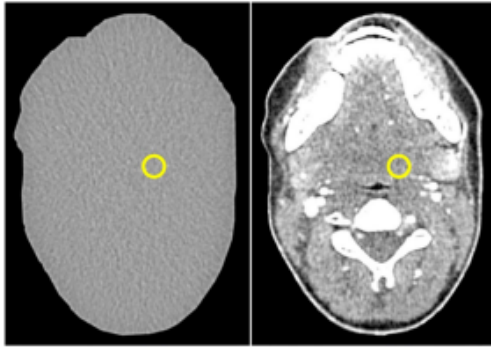
## Methods

### Study design

Neck-shaped phantoms with uniform and anatomical texture and hypodense lesions of 10 mm diameter and 4 to 38 HU contrast were imaged with two dose levels. Images were reconstructed with filtered back projection (FBP) and adaptive iterative dose reduction 3D (AIDR 3D). Lesion detectability was assessed by seven radiologists and compared between background types, dose levels, and reconstruction methods.

### Phantoms

Two phantom types, which were previously introduced for low-contrast detectability experiments, were used for this study: a uniform type consisting of polymethyl methacrylate with the shape of a patient's neck and a 3D printed, anatomically realistic type of identical shape [10, 11]. All phantoms had the same dimension of 15.4 cm (length) × 10.6 cm (width). Six different versions of the uniform phantom type and five versions of the anatomical type were used. One version of each type did not contain any lesion. The other versions each contained a single low-contrast lesion of 10 mm diameter in the left parapharyngeal space. The lesion was in the same position in all phantoms. Lesion contrasts were 4, 9, 18, 30, and 38 HU (uniform phantom) and 10, 18, 30, and 38 HU (anatomical phantom). The lesion contrasts were validated in previous studies by HU measurement in 2700 images acquired with six different dose levels (uniform phantom) and in 2808 images acquired with twenty-seven different dose levels (anatomical phantom) [10, 11]. In these validation experiments, lesion contrast was calculated as HU difference between regions of interest (ROIs) of 0.5 cm<sup>2</sup> inside the lesions and six ROIs of 4.9 cm<sup>2</sup> (uniform phantom) and one ROI of 3 cm<sup>2</sup> (anatomical phantom) surrounding the lesions. The lesions were rod-shaped, and the phantoms were constructed in such a way that multiple adjacent images displaying the same lesion and phantom background could be extracted per CT acquisition. Figure 1 shows a CT image of each phantom type and indicates the lesion position. Details on phantom construction, acquisitions, and measurements performed for evaluating lesion contrasts can be found elsewhere [10, 11].



**Fig. 1** CT images of the uniform and the anatomical phantom. The yellow circle indicates the lesion position

### Image acquisition

CT images of the uniform phantom originated from a previous study [11] and were acquired on a Canon Aquilion Prime CT scanner (Canon Medical Systems). CT images of the anatomical phantoms were acquired on the same system using identical parameters: helical mode, tube voltage of 120 kVp, fixed collimation of  $80 \times 0.5$  mm, rotation time of 0.5 s, 0.813 pitch, and a 280 mm diameter field of view. A 30- and 120-mA tube currents were used, corresponding to CTDIvol values of 1.4 and 5.6 mGy. Five acquisitions were performed per tube current. Images were reconstructed with 0.5-mm slice thickness and a soft tissue kernel (FC08) using FBP and AIDR 3D. For the subsequent detectability experiment, four CT images were extracted per acquisition of the lesion-bearing phantoms with 9, 18, 30, and 38 HU lesion contrast (uniform phantom) and 10, 18, 30, and 38 HU lesion contrast (anatomical phantom). Thus, a total of 640 lesion-bearing images were extracted (2 phantom types  $\times$  4 lesion contrasts  $\times$  2 tube currents  $\times$  2 reconstruction methods  $\times$  5 repeated acquisitions  $\times$  4 images).

### Detectability experiment

Each lesion-bearing image was paired with three non-lesion-bearing images of the corresponding phantom type (uniform or anatomical), which were acquired and reconstructed with identical parameters. Each of the resulting 640 image quartets was presented to seven radiologists in a 4-alternative forced choice (4-AFC) experiment. Readers were asked to select the image containing a lesion and to indicate their confidence using a five-step scale ranging from 1 = not confident to 5 = confident. Readings were performed using in-house developed software on diagnostic screens (Eizo RadiForce RX250, Eizo Corporation). In addition to the reading results obtained here,

results from a previous reading experiment performed with images of the uniform phantom and 4 HU lesion contrast were included in the analysis [11]. Image acquisitions and readings in that previous study were performed in the same way as in the present study (i.e., the same CT system, acquisition and reconstruction parameters, 4-AFC methodology, and readers were involved). The results were included to complement the current data used to analyze dose and image reconstruction effects in uniform phantoms.

### Noise characteristics

The standard deviation (SD) of pixel values and the noise power spectrum (NPS) were measured using 200 images per phantom type, tube current, and reconstruction method. In each image, a square ROI of  $32 \times 32$  pixels ( $17.5 \times 17.5$  mm) was placed in the same location in the parapharyngeal space adjacent to the lesion. The ROI position was selected to include a fairly homogeneous area of the anatomical phantoms. A larger ROI size or multiple ROIs would have led to the inclusion of largely inhomogeneous areas of the anatomical phantoms such as the mandibula or vascular structures. Also, ROI placement inside the lesions was not possible because the lesion size was too small to perform NPS measurement. The 2D NPS was calculated using the following Eq. (1):

$$NPS(f_x, f_y) = \frac{b_x b_y}{L_x L_y} \left\langle \left| \text{FFT}_{2D}[\text{ROI}(x, y) - \text{ROI}_{\text{Background}}(x, y)] \right|^2 \right\rangle_{N_{\text{ROI}}} \quad (1)$$

where  $b_x$  and  $b_y$  are the pixel sizes (0.546 mm) in the x- and y-direction, respectively, and  $L_x$  and  $L_y$  are the ROI lengths (17.5 mm) in the x- and y-direction, respectively.  $\text{FFT}_{2D}$  is the 2D fast Fourier transform.  $\text{ROI}_{\text{Background}}$  is the background noise in  $\text{ROI}(x, y)$  measured using second-order polynomial fitting by minimizing the residual sum of squares [12].  $N_{\text{ROI}}$  is the number of ROIs (200) per phantom type, tube current, and image reconstruction that was used to average the squared amplitude of the fast Fourier transform.

### Data and statistical analysis

Detection accuracy was calculated as the percentage of correct lesion image selections per reader. Detection accuracy and reader confidence were compared between uniform and anatomical phantom backgrounds using t-tests. Results were compared between dose levels and reconstruction methods with analysis of variance for repeated measurement using post hoc tests with Tukey's method to adjust for multiple comparisons. Differences were interpreted as significant for  $p < 0.05$ .

## Results

### Comparison of phantom types

Figure 2 shows a comparison of detection accuracy and reader confidence between uniform and anatomical phantoms. Averaged results across all readers, dose levels, and reconstruction methods are presented. Phantom background texture significantly affected detectability at all lesion contrasts. Readings of images of the uniform phantom yielded high detection accuracy already at relatively low lesion contrast of 9 HU (89.5%, 95% CI: 82.9 to 96%), which improved to 99.6% (95% CI: 99.1 to 100.2%) at 18 HU and perfect detection at 30 and 38 HU contrast. Conversely, readings of images of the anatomical phantoms yielded low detection accuracy at 10 HU (52.9%, 95% CI: 44.1 to 61.6%) and 18 HU (55.5%, 95% CI: 47.2 to 63.9%), which improved to 91.1% (95% CI: 85.8 to 96.3%) at 30 HU and 97.5% (95% CI: 95.8 to 99.2%) at 38 HU contrast. Clear differences between uniform and anatomical images were also observed for reader confidence (Fig. 2, suppl. table 1). Similar detection accuracies for the two phantom types were achieved when comparing 9 HU lesion contrast in the uniform phantom and 30 HU contrast in the anatomical phantom (89.5% vs. 91.1%,  $p=0.587$ ). Readings of images of the uniform phantom with 4 HU lesion contrast originating from a previous study yielded an average detection accuracy of 62.9%

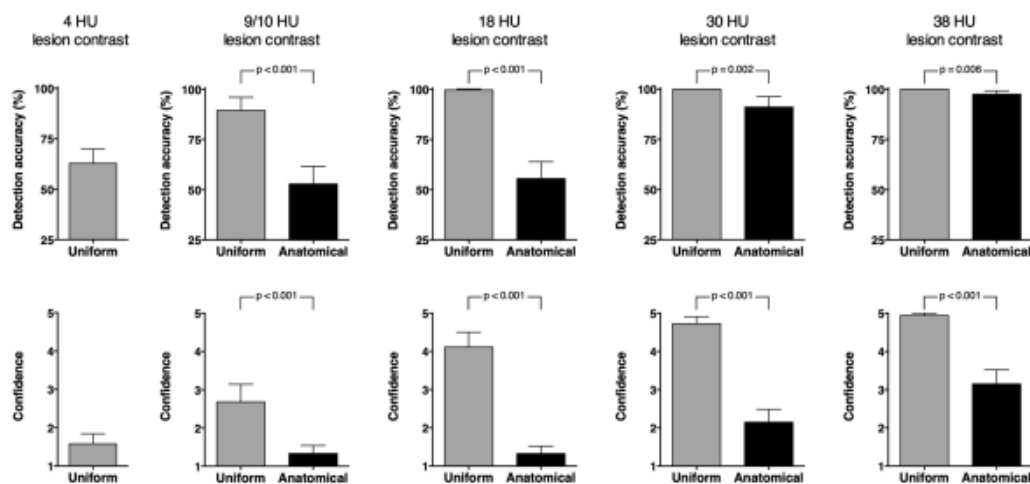
across all readers, dose levels, and reconstruction methods (95% CI: 55.8 to 69.9%) [11].

### Comparison of dose and image reconstruction

Figure 3 provides a series of uniform and anatomical phantom images acquired at 1.4 and 5.6 mGy and reconstructed with FBP and AIDR 3D. The figure includes uniform images with 9 HU lesion contrast and anatomical images with 30 HU lesion contrast, which yielded similar overall detection accuracies. Detailed detection accuracy results per dose, reconstruction method, and lesion contrast are presented in Tables 1 and 2.

At 18 HU lesion contrast and above, readings of images of the uniform phantom reached 100% detection accuracy and could therefore not be used for the analysis of dose and image reconstruction effects. Results for 4 and 9 HU lesion contrast are summarized in Table 3 and presented in Fig. 4. Dose reduction from 5.6 to 1.4 mGy decreased lesion detectability in uniform images that were reconstructed with FBP (83.6% vs. 55.4%,  $p < 0.001$ ). AIDR 3D maintained detectability (85% vs. 80.7%,  $p = 0.375$ ) and was superior to FBP at 1.4 mGy ( $p < 0.001$ ). Analysis of the uniform phantom thus showed strong dose effects on FBP-reconstructed images and superiority of AIDR 3D at 1.4 mGy.

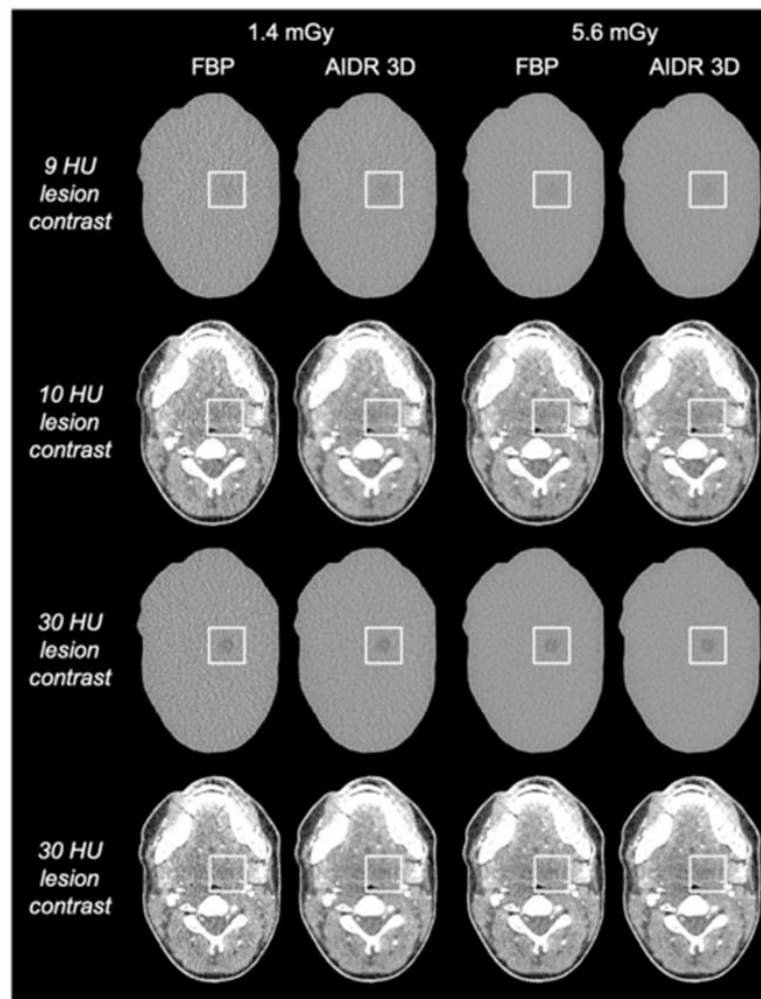
Figure 5 shows the effects of dose and image reconstruction on detection in anatomical phantoms. Numerical results are provided in Table 4. In contrast to the uniform phantom,



**Fig. 2** Detection accuracy and reader confidence in uniform and anatomical phantom images. Averaged results across all readers, dose levels, and reconstruction methods at 4, 9 (uniform), 10 (anatomical),

18, 30, and 38 HU lesion contrast are presented. Error bars indicate 95% confidence intervals. 25% detection accuracy corresponds to random guessing (no detection)

**Fig. 3** Uniform and anatomical phantom images acquired at 1.4 and 5.6 mGy and reconstructed with filtered back projection (FBP) and adaptive iterative dose reduction 3D (AIDR 3D)



**Table 1** Detection accuracy (%) per dose, reconstruction method, and lesion contrast in uniform phantom images. Means and 95% confidence intervals are presented

		4 HU lesion contrast	9 HU lesion contrast	18 HU lesion contrast	30 HU lesion contrast	38 HU lesion contrast
1.4 mGy	FBP	42.9 (27.3 to 58.4)	67.9 (48.8 to 86.9)	99.3 (97.5 to 101)	100	100
	AIDR 3D	67.9 (57.9 to 77.8)	93.6 (84.4 to 102.7)	100	100	100
5.6 mGy	FBP	70 (55.9 to 84.1)	97.1 (93.5 to 100.8)	99.3 (97.5 to 101)	100	100
	AIDR 3D	70.7 (56.7 to 84.7)	99.3 (97.5 to 101)	100	100	100

**Table 2** Detection accuracy (%) per dose, reconstruction method, and lesion contrast in anatomical phantom images. Means and 95% confidence intervals are presented

		10 HU lesion contrast	18 HU lesion contrast	30 HU lesion contrast	38 HU lesion contrast
1.4 mGy	FBP	48.6 (26.3 to 70.9)	45 (23 to 67)	85.7 (66.2 to 105.2)	93.6 (89.2 to 98)
	AIDR 3D	47.9 (25.5 to 70.2)	54.3 (36.5 to 72.1)	90.7 (79.3 to 102.2)	99.3 (97.5 to 101)
5.6 mGy	FBP	48.6 (26.3 to 70.9)	56.4 (41.6 to 71.2)	95.7 (89 to 102.5)	97.9 (92.6 to 103.1)
	AIDR 3D	66.4 (51.4 to 81.5)	66.4 (43.2 to 89.7)	92.1 (82.2 to 102.1)	99.3 (97.5 to 101)

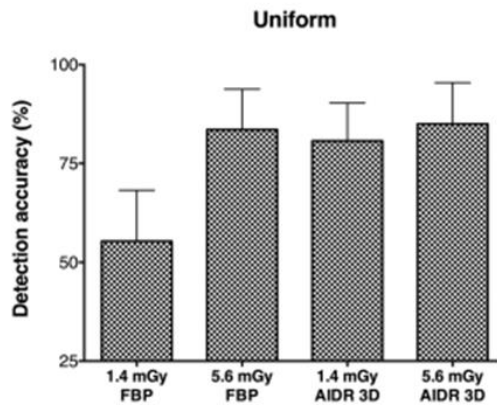
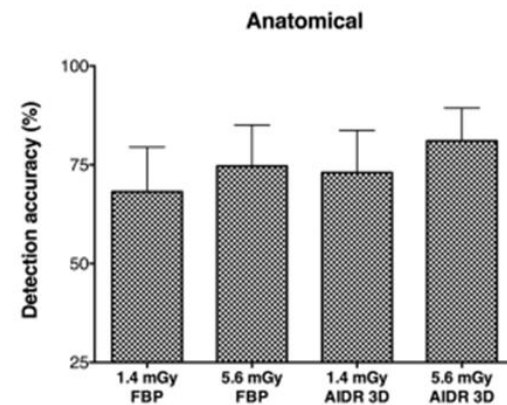
**Table 3** Detection accuracy (%) per dose and reconstruction method in uniform phantom images. Averaged results across 4 and 9 HU lesion contrast and 95% confidence intervals are presented

	FBP	AIDR 3D	<i>p</i> value
1.4 mGy	55.4 (42.5 to 68.2)	80.7 (71.1 to 90.3)	< 0.001
5.6 mGy	83.6 (73.4 to 93.8)	85 (74.6 to 95.4)	0.785
<i>p</i> value	< 0.001	0.375	

AIDR 3D did not show clear advantages over FBP at any dose level (73% vs. 68.2%,  $p=0.144$  at 1.4 mGy and 81.1% vs. 74.6%,  $p=0.111$  at 5.6 mGy). Moreover, the strong effects of dose reduction on FBP-reconstructed images were not confirmed. Instead, dose reduction moderately affected detectability in a similar manner for both reconstruction methods ( $p=0.027$  for FBP and  $p=0.018$  for AIDR 3D). Analysis of the anatomical phantoms thus neither confirmed the superiority of AIDR 3D nor dose effects observed in the uniform phantom.

### Noise characteristics

Figure 6 shows noise and NPS results per phantom type, dose, and image reconstruction. Numerical results are summarized in Table 5. As expected, low-dose (1.4 mGy), FBP-reconstructed images had the highest noise level in both phantom types ( $p < 0.001$ ). A dose increase to 5.6 mGy reduced the noise ( $p < 0.001$ ) except for AIDR 3D-reconstructed images of the anatomical phantoms, which had almost identical noise values at low and high doses ( $p=0.26$ ). Remarkably, noise was lower in low-dose AIDR 3D-reconstructed images than in high-dose FBP-reconstructed images of the anatomical, but not of the uniform, phantom, indicating that AIDR 3D was more effective in denoising anatomical images. The NPS curves of the uniform phantom showed a shift towards lower spatial frequencies in low-dose AIDR 3D-reconstructed images with a peak NPS at  $0.23 \text{ mm}^{-1}$  and a decrease at lower spatial frequencies. Conversely, all images of the anatomical phantoms yielded peak NPS values at a low spatial frequency of  $0.12 \text{ mm}^{-1}$  regardless of dose and image reconstruction. FBP-reconstructed images acquired at 1.4 mGy had a second

**Fig. 4** Detection accuracy per dose and reconstruction method in uniform phantom images. Averaged results across 4 and 9 HU lesion contrast are presented. Error bars indicate 95% confidence intervals. 25% detection accuracy corresponds to random guessing (no detection)**Fig. 5** Detection accuracy per dose and reconstruction method in anatomical phantom images. Averaged results across all lesion contrasts are presented. Error bars indicate 95% confidence intervals. 25% detection accuracy corresponds to random guessing (no detection)

**Table 4** Detection accuracy (%) per dose and reconstruction method in anatomical phantom images. Averaged results across all lesion contrasts and 95% confidence intervals are presented

	FBP	AIDR 3D	<i>p</i> value
1.4 mGy	68.2 (56.9 to 79.5)	73 (62.3 to 83.7)	0.144
5.6 mGy	74.6 (64.2 to 85.1)	81.1 (72.8 to 89.3)	0.111
<i>p</i> value	0.027	0.018	

NPS peak at a spatial frequency of  $0.23 \text{ mm}^{-1}$ , which flattened with FBP reconstruction at 5.6 mGy and in all images reconstructed with AIDR 3D.

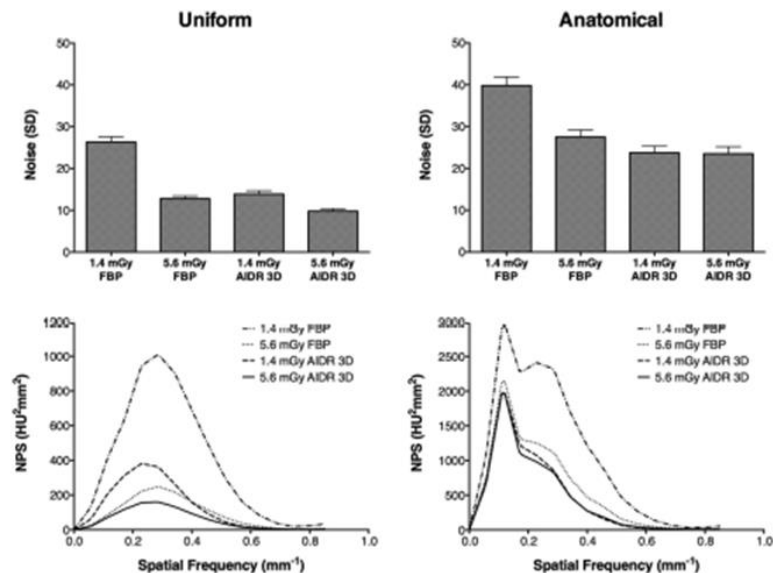
## Discussion

Task-based methods have been proposed to evaluate and compare CT techniques for their diagnostic performance in clinical practice. Task-based assessment is typically performed using CT images of uniform phantoms, and it is of interest to what extent evidence from uniform phantoms actually reflects detectability in clinical images with anatomical detail. The present study therefore compared low-contrast detectability between uniform and anatomically realistic phantoms. Our results show that anatomical phantom structure affects detection accuracy at all investigated lesion contrasts ( $p < 0.001$ ), interferes with dose effects on detection and influences the assessment of AIDR 3D performance compared to FBP.

The image assessment results we obtained for the uniform phantom are in good agreement with previous reports of relatively high detection sensitivities of more than 87% for lesions of the same size as investigated in our study [13, 14]. Anatomical phantom structure significantly impaired lesion detectability—a contrast increase to 30 HU was necessary to achieve similar detection accuracy as for 9 HU lesion contrast in uniform images. Near-perfect detectability was achieved at a markedly higher lesion contrast (38 HU) than with the uniform phantom (18 HU).

An impact of anatomical detail was expected because structured tissue patterns (anatomical noise) have psychophysical effects on humans that interfere with detection tasks. Previous X-ray studies found anatomical noise to have stronger effects than quantum noise and to impair and eventually limit human lesion perception [5–7]. This, in turn, may influence how dose changes affect detection tasks [6, 7]. Our experiments confirm the effects of anatomical patterns on noise characteristics and the assessment of dose and reconstruction methods. Anatomical images had a low-frequency noise component that was predominant regardless of dose and image reconstruction mode. This component was in good agreement with reports of high NPS values at low spatial frequencies in patients [15]. Anatomical background structure also influenced the denoising power of AIDR 3D, which adds to reports on interactions between anatomical texture, noise, and spatial resolution when iterative reconstruction is applied [16–18]. Lesion detectability was clearly affected by dose in uniform FBP images. However, the dose-detection relationship was less clear in images

**Fig. 6** Noise and noise power spectrum (NPS) results. Average noise values from 200 images per phantom type, dose, and image reconstruction are presented. Error bars indicate standard deviations



**Table 5** Noise (SD) and spatial frequency ( $\text{mm}^{-1}$ ) of the peak amplitude of the noise power spectrum (NPS). Averaged noise results from 200 images per phantom type, dose, and image reconstruction are presented along with 95% confidence intervals

	1.4 mGy FBP	1.4 mGy AIDR 3D	5.6 mGy FBP	5.6 mGy AIDR 3D
Noise (SD)	26.33	13.89	12.79	9.8
Uniform	(26.15 to 26.51)	(13.77 to 14)	(12.7 to 12.88)	(9.72 to 9.88)
Noise (SD)	39.79	23.76	27.52	23.5
Anatomical	(39.5 to 40.08)	(23.54 to 23.98)	(27.28 to 27.75)	(23.27 to 23.74)
Spatial frequency of peak NPS ( $\text{mm}^{-1}$ )	0.28	0.23	0.29	0.28
Uniform				
Spatial frequency of peak NPS ( $\text{mm}^{-1}$ )	0.12	0.12	0.12	0.12
Anatomical				

with anatomical noise. Consistent with published results, AIDR 3D maintained detectability and was superior to FBP at a lower dose in uniform phantom images [19]. These advantages were lost when anatomical structures interfered with lesion detection.

Significant texture effects on detectability were also observed in a previous CT study that compared liver-mimicking textures with a uniform phantom background [8]. In that study, structured background textures reduced the influence of dose changes on detection, similar to what we observed for FBP images. Another CT study came to different conclusions and reported only negligible effects of liver texture on detectability in comparison with a water background [9]. However, liver and water textures in that study were visually quite similar, which explains why the results differ from our observations. However, it should also be noted that the comparability of our results with both of these CT studies is limited by differences in CT hardware and because both studies investigated cropped images with vessel-free liver textures. To the best of our knowledge, our study is the first to compare neck phantom images with full anatomical detail, which is relevant because anatomical detail adds complexity to CT images and has a relevant impact on human lesion perception [20, 21].

The experiments we performed here do not provide an in-depth analysis of dose reduction and image reconstruction, which requires broader testing and can be found elsewhere [22]. For example, AIDR 3D was reported to have similar performance as FBP at 120 kVp, which our experiments confirmed, and also to be superior at a lower tube voltage of 100 kVp, which we did not assess [22]. Our study evaluated the effects of phantom background on task-based CT assessment, and we used two dose levels and reconstructions methods to illustrate such effects. Based on our results, we conclude that phantom background has a relevant influence and that transferability of CT assessment to clinical imaging can be expected to improve as the realism of the test environment increases. In view of the published evidence discussed above, we believe that this should apply beyond the CT scanner and imaging technologies used here.

The limitations of our study include the rather narrow study protocol, which was selected to investigate the effects of phantom background, but not to perform a comprehensive analysis of dose and image reconstruction methods. Results may differ in less complex anatomical regions than the neck. However, the generalizability of our results is supported by previous work in liver imaging, which has arrived at similar conclusions about the importance of phantom texture [8]. It should also be noted that we deliberately chose a location-known-exactly experimental design in order to avoid introducing different lesion locations as another variable possibly influencing detectability. Yet, detection experiments with lesions in unknown locations can be considered to be more realistic and representative of clinical image interpretation [4].

Uniform phantoms differ from patients and provide an idealized environment for evaluating CT systems. Our results provide evidence that lesion contrasts in CT images of uniform phantoms are below those that are clinically relevant and corroborate data indicating that anatomical phantom structure affects estimates of CT performance and reasonable dose selection. Investigations of CT assessment aimed at predicting and comparing clinical performance must take into account differences between phantoms and patients and should be performed in a setting that mimics clinical imaging as closely as possible.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-021-08248-3>.

**Acknowledgements** We thank our colleagues from the Department of Radiology and Neuroradiology for participating in the detectability experiment, and Bettina Herwig for assistance with the preparation of the article. Dr. Jahnke is a participant in the BIH-Charité Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This study has received funding by the Bundesministerium für Wirtschaft und Energie (DE): 03EFHBE093.



## Declarations

**Guarantor** The scientific guarantor of this publication is Dr. Paul Jahnke.

**Conflict of interest** Dr. Jahnke and PD Dr. Scheel are patent inventors (EP3135199A1, US9924919B2, US10182786B2). Dr. Jahnke, PD Dr. Scheel and Prof. Dr. Hamm are shareholders of PhantomX GmbH.

**Statistics and biometry** No complex statistical methods were necessary for this paper.

**Informed consent** Written informed consent was waived by the Institutional Review Board.

**Ethical approval** Institutional Review Board approval was obtained.

**Study subjects or cohorts overlap** Some study subjects or cohorts have been previously reported in Conzelmann J, Schwarz FB, Hamm B, Scheel M, Jahnke P (2020) Development of a method to create uniform phantoms for task-based assessment of CT image quality. *J Appl Clin Med Phys* 21:201–208.

## Methodology

- prospective
- observational
- performed at one institution

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Racine D, Ryckx N, Ba A et al (2018) Task-based quantification of image quality using a model observer in abdominal CT: a multicentre study. *Eur Radiol* 28:5203–5210
2. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ (2014) Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Med Phys* 41:071904
3. Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP (2015) Task-based measures of image quality and their relation to radiation dose and patient risk. *Phys Med Biol* 60:R1–75
4. Samei E, Bakalyar D, Boedeker KL et al (2019) Performance evaluation of computed tomography systems: summary of AAPM Task Group 233. *Med Phys* 46:e735–e756
5. Samei E, Flynn MJ, Eyler WR (1999) Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs. *Radiology* 213:727–734
6. Kotre CJ (1998) The effect of background structure on the detection of low contrast objects in mammography. *Br J Radiol* 71:1162–1167
7. Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P (1999) Estimation of the noisy component of anatomical backgrounds. *Med Phys* 26:1365–1370
8. Solomon J, Ba A, Bochud F, Samei E (2016) Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms. *Med Phys* 43:6497
9. Dilger SKN, Yu L, Chen B et al (2019) Localization of liver lesions in abdominal CT imaging: I. Correlation of human observer performance between anatomical and uniform backgrounds. *Phys Med Biol* 64:105011
10. Ardila Pardo GL, Conzelmann J, Genske U, Hamm B, Scheel M, Jahnke P (2020) 3D printing of anatomically realistic phantoms with detection tasks to assess the diagnostic performance of CT images. *Eur Radiol* 30:4557–4563
11. Conzelmann J, Schwarz FB, Hamm B, Scheel M, Jahnke P (2020) Development of a method to create uniform phantoms for task-based assessment of CT image quality. *J Appl Clin Med Phys*. <https://doi.org/10.1002/acm2.12974>
12. Dolly S, Chen HC, Anastasio M, Mutic S, Li H (2016) Practical considerations for noise power spectra estimation for clinical CT scanners. *J Appl Clin Med Phys* 17:392–407
13. Euler A, Stieltjes B, Szucs-Farkas Z et al (2017) Impact of model-based iterative reconstruction on low-contrast lesion detection and image quality in abdominal CT: a 12-reader-based comparative phantom study with filtered back projection at different tube voltages. *Eur Radiol* 27:5252–5259
14. Schindera ST, Oedra D, Raza SA et al (2013) Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology* 269:511–518
15. Ba A, Abbey CK, Racine D et al (2019) Channelized Hotelling observer correlation with human observers for low-contrast detection in liver CT images. *J Med Imaging (Bellingham)* 6:025501
16. Solomon J, Samei E (2014) Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. *Med Phys* 41:091908
17. 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* 39:4115–4122
18. Yu L, Vrieze TJ, Leng S, Fletcher JG, McCollough CH (2015) Technical note: measuring contrast- and noise-dependent spatial resolution of an iterative reconstruction method in CT using ensemble averaging. *Med Phys* 42:2261–2267
19. Joemai RM, Veldkamp WJ, Kroft LJ, Hernandez-Giron I, Geleijns J (2013) Adaptive iterative dose reduction 3D versus filtered back projection in CT: evaluation of image quality. *AJR Am J Roentgenol* 201:1291–1297
20. Samei E, Flynn MJ, Peterson E, Eyler WR (2003) Subtle lung nodules: influence of local anatomic variations on detection. *Radiology* 228:76–84
21. Smith TB, Rubin GD, Solomon J, Harrawood B, Choudhury KR, Samei E (2018) Local complexity metrics to quantify the effect of anatomical noise on detectability of lung nodules in chest CT imaging. *J Med Imaging (Bellingham)* 5:045502
22. Jahnke P, Conzelmann J, Genske U et al (2021) Task-based assessment of neck CT protocols using patient-mimicking phantoms—effects of protocol parameters on dose and diagnostic performance. *Eur Radiol* 31:3177–3186

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

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

## Komplette Publikationsliste

### Erstautorenschaften:

1. **Comparison of low-contrast detectability between uniform and anatomically realistic phantoms-influences on CT image quality assessment.**  
Juliane Conzelmann, Ulrich Genske, Arthur Emig, Michael Scheel, Bernd Hamm, Paul Jahnke. Eur Radiol. 2022 Feb;32(2):1267-1275. doi: 10.1007/s00330-021-08248-3.  
Impact Factor 2021: 7,034
2. **Development of a method to create uniform phantoms for task-based assessment of CT image quality**  
Juliane Conzelmann, Felix Benjamin Schwarz, Bernd Hamm, Michael Scheel, Paul Jahnke. J Appl Clin Med Phys. 2020 Sep;21(9):201-208. doi: 10.1002/acm2.12974.  
Impact Factor 2020: 2,102

### Coautorenschaften:

1. **Task-based assessment of neck CT protocols using patient-mimicking phantoms-effects of protocol parameters on dose and diagnostic performance**  
Paul Jahnke, Juliane Conzelmann, Ulrich Genske, Maximilian Nunninger, Michael Scheel, Bernd Hamm, Torsten Diekhoff. Eur Radiol. 2021 May;31(5):3177-3186. doi: 10.1007/s00330-020-07374-8.  
Impact Factor 2021: 7,034
2. **3D printing of anatomically realistic phantoms with detection tasks to assess the diagnostic performance of CT images.**  
Gracia Lana Ardila Pardo, Juliane Conzelmann, Ulrich Genske, Bernd Hamm, Michael Scheel, Paul Jahnke. Eur Radiol. 2020 Aug;30(8):4557-4563. doi: 10.1007/s00330-020-06808-7.  
Impact Factor 2020: 5,315

## **Danksagung**

Als allererstes möchte ich mich bei meinen Betreuern PD Dr. Paul Jahnke und PD Dr. Michael Scheel dafür bedanken, dass sie immer ansprechbar waren, mich in jeder Phase der Doktorarbeit unterstützt und das Projekt ermöglicht haben.

Außerdem möchte ich mich bei dem ganzen Team der AG Jahnke für die herzliche Aufnahme in die Arbeitsgruppe bedanken, sowie bei allen Radiologen und Radiologinnen, die mit viel Zeiteinsatz an den Studien teilgenommen haben.

Besonderer Dank gilt meine Schwester Johanna Buttenberg und ihrer Familie sowie meinem Freund Camilo Martinez Barón für die fürsorgliche Unterstützung.

Zuletzt möchte ich mich bei meinen Eltern für ihren unerschöpflichen Rückhalt bedanken.