

Aus der

Klinik und Poliklinik für Radiologie  
der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. med. Jens Ricke

**Anwendung von Künstlicher Intelligenz und Strukturierter Befundung in der  
Radiologischen Diagnostik**

**Habilitationsschrift**

zur Erlangung der Venia Legendi  
für das Fach Radiologie  
der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität München

Vorgelegt von

Dr. med. Bastian Sabel

2021

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitende Zusammenfassung.....</b>	<b>3</b>
<b>2</b>	<b>Themenrelevante wissenschaftliche Arbeiten.....</b>	<b>7</b>
2.1	Strukturierte Befundung der CT-Pulmonalisangiographie bei akuter Lungenembolie.....	7
2.2	Strukturierte Befundung der CT-Angiographie der unteren Extremitäten.....	9
2.3	KI-Unterstützung zur Befundung von CT-Kontrollen thorakaler Aortenaneurysmen.....	11
2.4	KI-unterstützte Befundung zur Reduzierung von übersehenen Pathologien in notfallmäßig durchgeführten Schockraum-CT Untersuchungen.....	14
2.5	KI-Algorithmus zur Erkennung von Pneumonien in Thorax-Röntgenbildern.....	15
2.6	Bedeutung von Thoraxdrainagen und Ausmaß der Pleuradehiszenz für die Pneumothorax-Erkennung im Röntgenthorax durch KI-Algorithmen.....	18
2.7	Prognostischer Wert des CT-Thorax für die invasive Beatmungstherapie bei COVID-19-Pneumonien.....	21
2.8	Differenzierung zwischen COVID-19, Influenza und anderen Pneumonien mittels CT.....	24
<b>3</b>	<b>Literaturverzeichnis.....</b>	<b>26</b>

# 1 Einleitende Zusammenfassung

Verglichen mit den bedeutenden Fortschritten der Radiologie in der Bildgebung selbst, blieb die Art und Weise der radiologischen Befunderstellung innerhalb der letzten Jahrzehnte weitgehend unverändert. Erst in den letzten Jahren gab es Anstrengungen, die Form der in Prosa gehaltenen Befundberichte in eine klare und offensichtliche Strukturierung zu überführen, um den klinischen Nutzen der radiologischen Diagnostik zu erhöhen. „Structured Reporting“ wurde unter anderem von der Radiological Society of North America (RSNA), der European Society of Radiology (ESR) und der Deutschen Röntgengesellschaft (DRG) unterstützt und gefördert. Strukturierte Befundberichte enthalten eine konsistente Anordnung der Beobachtungen (52). Eine Vorlage mit standardisierten Überschriften und modularem Format macht es für den Radiologen unwahrscheinlicher, wichtige Deskriptoren zu übersehen (47). Studien haben gezeigt, dass Strukturierte radiologische Befundberichte mit standardisierter Klassifikation sowohl effizienter und vollständiger sind als auch Fehlerquellen minimieren (19, 27). Insbesondere bei weniger erfahrenen Radiologen stellen sie die Vollständigkeit der Befundberichte sicher (8). Strukturierte Befundberichte sind Konventionellen Befundberichten in Genauigkeit und Verständlichkeit überlegen und verbessern die Klarheit der Kommunikation zwischen dem Radiologen und dem zuweisenden Arzt (6, 46). Außerdem lassen sich die für die Behandlung des Patienten notwendigen Informationen leichter extrahieren (5, 15, 34). Folglich kann eine Strukturierte Befundung auch eine deutliche therapeutische Konsequenz für betroffene Patienten haben, die mit einem besseren Outcome verbunden sein kann (5).

**Teilprojekt 1** evaluierte die Strukturierte Befundung von CT-Pulmonalisangiographie (CTPA) Untersuchungen hinsichtlich Inhalt, Klarheit und klinischem Nutzen bei akuter Lungenembolie (LAE). Es konnte gezeigt werden, dass die zuweisenden Fachkliniker Strukturierte CTPA-Befundberichte gegenüber Konventionellen Befundberichten in ihrer Klarheit als überlegen sehen. Ebenso schätzen Pulmologen Strukturierte Befundberichte in Bezug auf Inhalt und klinischen Nutzen gegenüber Konventionellen Befundberichten als überlegen ein. Eine Änderung im Patientenmanagement scheint Strukturierte Befundung bei akuter LAE nicht zu bewirken.

Ziel von **Teilprojekt 2** war es anschließend, die Auswirkung einer Strukturierten Befundung von Computertomographie-Angiographie (CTA) Untersuchungen bei Patienten mit bekannter oder vermuteter peripherer arterieller Verschlusskrankheit auf ihre Klarheit, Vollständigkeit, klinische Relevanz und Nützlichkeit zu untersuchen. Darüber hinaus wurden die sich aus den jeweiligen Befundtexten ergebenden weiterführenden diagnostischen Untersuchungen und Therapiekonzepte betrachtet. Die Ergebnisse zeigten, dass zuweisende Kliniker Strukturierten Befundberichte von CTA Untersuchungen der unteren Extremitäten in Bezug auf Klarheit, Vollständigkeit, klinische Relevanz und Nützlichkeit gegenüber Konventionellen Befundberichten als überlegen bewerten. Die Strukturierte Befundung scheint die weitere Untersuchung oder Therapie bei Patienten mit bekannter oder vermuteter peripherer arterieller Verschlusskrankheit nicht zu beeinflussen.

Aufgrund der nachgewiesenen Vorteile der Strukturierten Befundung gehen wir davon aus, dass diese langfristig den klinischen Nutzen der Befundberichte in allen Teilbereichen der Radiologie erhöhen wird. Geradezu nebenbei eignen sich die hierbei generierten Befundberichte deutlich besser als Freitextbefunde für das gezielte Training von Algorithmen der Künstlichen Intelligenz (KI). Aufgrund seiner bereits fortgeschrittenen Digitalisierung, welche bereits vor vielen Jahren mit der Einführung von Computer Aided Diagnostic (CAD) Systemen begonnen hat, ist damit zu rechnen, dass die Radiologie eines der ersten medizinischen Fächer mit relevanter Bedeutung von KI in der täglichen Routine sein wird. Die Teilprojekte 3 bis 6 beschäftigen sich mit beispielhaften Anwendungen von KI in der Radiologie und beleuchten dabei auf der einen Seite deren Potential zur Steigerung von Qualität, Zuverlässigkeit und Effizienz in der radiologischen Befundung und auf der anderen Seite gefährliche Stolpersteine bei der Entwicklung und Etablierung von KI-Algorithmen.

**Teilprojekt 3** befasst sich mit dem Einsatz von Künstlicher Intelligenz für die quantitative Beurteilung von Aortenaneurysmen. Die leitlinienkonforme Aortenvermessung ist von entscheidender Bedeutung für die präzise Erfassung der Aneurysma-Ausdehnung und kann die Behandlungsstrategie stark beeinflussen. Die Ergebnisse von Teilprojekt 3 zeigen, dass sich durch KI-Unterstützung zur Aortenvermessung sowohl ein Zeitersparnis als auch eine Verringerung der Inter-Reader-Variabilität von Radiologen erreichen lassen.

Die radiologische Befundung von notfallmäßigen Schockraum-CT Untersuchungen ist zeitkritisch und birgt daher ein erhebliches Risiko, dass insbesondere Pathologien abseits der akuten Fragestellung übersehen werden können. In **Teilprojekt 4** zeigten wir, dass mithilfe eines KI-gestützten Softwareprototypen, bestehend aus einzelnen pathologiespezifischen AI-Algorithmen, die Anzahl an übersehenen relevanten Nebenbefunden in Schockraum-CT Untersuchungen reduziert werden kann. Die hohe Anzahl an falsch-positiven oder nicht relevanten KI-Detektionen machte deutlich, dass die Algorithmen für ihre Anwendung im klinischen Alltag hinsichtlich ihrer Spezifität optimiert werden sollten.

Der Nutzen eines KI-Algorithmus für die Beurteilung von pulmonalen Verdichtungen in Liegend-Röntgenthoraxaufnahmen wurde in **Teilprojekt 5** behandelt. Es konnte gezeigt werden, dass ein KI-Algorithmus, in Anbetracht der geringen Performance-Unterschiede zwischen dem KI-Algorithmus und den Radiologen, als vielversprechende klinische Entscheidungshilfe für Thorax-Röntgenuntersuchungen in der klinischen Routine gesehen werden kann. Der Algorithmus offenbarte großes Potenzial für eine Reduktion der Rate an übersehenen Pathologien durch eine KI-Unterstützung des Radiologen.

**Teilprojekt 6** offenbart schließlich, dass bei der Entwicklung und Nutzung von KI-Algorithmen eine sorgfältige Qualitätssicherung erforderlich ist. Wir stellten in diesem Teilprojekt die Hypothese auf, dass die publizierten Ergebnisse von etablierten Algorithmen zur Erkennung von Pneumothoraces in Thorax-Röntgenbildern den Einfluss der Pneumothorax-Größe und störende Effekte durch Thoraxdrainagen nicht ausreichend berücksichtigen. Unsere detaillierte Subgruppenanalyse hat gezeigt, dass die Leistungsfähigkeit von auf öffentlichen Datensätzen trainierten KI-Algorithmen zur Pneumothorax-Erkennung stark von der Pneumothorax-Größe abhängt und durch störende Bildmerkmale, wie z. B. eingelegte Thoraxdrainagen, signifikant verzerrt wird.

Insbesondere in Situationen knapper Ressourcen oder wenn folgenschwere Entscheidungen schnell getroffen werden müssen, ist eine eindeutige und effiziente Kommunikation von großer Bedeutung. Die weltweite COVID-19 Pandemie forderte gerade in ihrer frühen Phase von Radiologen ab, sich in Röntgen-Bildern und CT-Untersuchungen zügig und verlässlich hinsichtlich COVID-19 typischer oder untypischer Befunde festzulegen. Aufgrund der begrenzten Kapazitäten in den Notaufnahmen der Krankenhäuser und der langen Latenz bis

zum Erhalt eines PCR-Tests mussten Patientendisposition und Therapien nicht selten auf Grundlage des radiologischen Befundes entschieden werden.

In **Teilprojekt 7** sollten Thorax-CT-Bildgebungsmerkmale der COVID-19-Erkrankung bei Krankenhausaufnahme zur Risikostratifizierung für invasive Beatmung (IV) oder keine bzw. nicht-invasive Beatmung (non-IV) während des Krankenhausaufenthalts bewertet werden. Das von uns erarbeitete Scoring-Modell bietet ein zuverlässiges und effektives Werkzeug zur Erstbeurteilung von COVID-19-Patienten und unterstützt die effektive Zuweisung und das Ressourcenmanagement innerhalb von Krankenhäusern und Krankenhausnetzwerken.

Die zeitaufwendige real-time Polymerase-Kettenreaktion (RT-PCR) des Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) leidet in frühen Infektionsstadien unter einer begrenzten Sensitivität, während ein schnell verfügbares Thorax-CT bereits einen COVID-19-Verdacht begründen kann. Bisher war die diagnostische Bedeutung von Radiologen hinsichtlich der Differenzierung von COVID-19 zu anderen atypischen Pneumonien, insbesondere der Influenza-Pneumonie, nicht ausreichend belegt. In **Teilprojekt 8** bestätigten wir die diagnostische Aussagekraft der Thorax-CT-Beurteilung durch erfahrene Radiologen zur sensitiven Identifizierung einer COVID-19-Pneumonie im frühen Infektionsstadium. Diese wurde im Rahmen dieser Arbeit auch mit der zur Charakterisierung von anderen atypischen Pneumonien, insbesondere der Influenza-Infektion, verglichen.

## 2 Themenrelevante wissenschaftliche Arbeiten

### 2.1 Strukturierte Befundung der CT-Pulmonalisangiographie bei akuter Lungenembolie

**Teilprojekt 1: Sabel BO\***, Plum JL\*, Kneidinger N, Leuschner G, Koletzko L, Raziorrouh B, Schinner R, Kunz WG, Schoeppe F, Thierfelder KM, Sommer WH, Meinel FG. Structured reporting of CT examinations in acute pulmonary embolism. **J Cardiovasc Comput Tomogr.** doi: 10.1016/j.jcct.2017.02.008. \*geteilte Erstautorenschaft

*Journal Impact Factor: 2,9*

Die Strukturierte Befundung fördert durch ihre Werkzeuge zur Leitlinien-konformen Dokumentation (Checklisten, klinische Entscheidungshilfen) die eindeutige, vollständige und konsistente Kommunikation von radiologischen Befundberichten. Aktuelle Bestrebungen der Radiologie zu einer einheitlichen Befundungsnomenklatur (z. B. die RadLex-Initiative) und zu modernen IT-Systemen unterstützen diese Bestrebungen zusätzlich. Für CT-Untersuchungen bei Patienten mit Lungenembolie wurde der potenzielle Mehrwert einer standardisierten Strukturierten Befundung bisher nicht untersucht. Ziel von Teilprojekt 1 war es daher, den Effekt einer Strukturierten Befundung von CT-Pulmonalangiographie (CTPA)-Untersuchungen auf den Inhalt, die Eindeutigkeit und den klinischen Nutzen der Befundberichte bei akuter Lungenembolie (LAE) zu evaluieren.

In dieser Arbeit wurden jeweils Konventionelle und Strukturierte Befundberichte für 41 Patienten mit CTPA und positivem Nachweis einer LAE erstellt. Die Konventionellen Befundberichte wurden in Freitextform mittels Spracherkennung diktiert; die Strukturierten Befundberichte enthielten eine konsistente Anordnung der Befundbeschreibungen mit standardisierten Zwischenüberschriften. Die Konventionellen Befundberichte wurden mit den standardisierten Strukturierten Befundberichten verglichen. Zwei allgemeine Internisten und zwei spezialisierte Pulmologen bewerteten ihre Zufriedenheit mit den Befundberichten hinsichtlich Eindeutigkeit und Inhalt sowie klinischem Nutzen und ordneten die Patienten hypothetisch einer geeigneten Zuständigkeit (Ambulanz, Normalstation, Intensivstation) und Therapie (einschließlich systemischer Thrombolyse, interventioneller oder chirurgischer Thrombektomie) zu.

Die Strukturierten Befundberichte wurden im Vergleich zu den Konventionellen Befundberichten bezüglich ihres Inhalts (mediane Bewertung 10,0 vs. 8,5,  $p < 0,0001$ ) und ihrer Eindeutigkeit (mediane Bewertung 10,0 vs. 8,0) höher bewertet. Der Anstieg der Zufriedenheit hinsichtlich Eindeutigkeit war sowohl für allgemeine Internisten als auch für spezialisierte Pulmologen signifikant ( $p < 0,0001$ ), während sich die Zufriedenheit mit dem Inhalt nur bei Pulmologen als signifikant höher erwies ( $p < 0,0001$ ). Der klinische Nutzen der radiologischen Befunde stieg mit der Strukturierten Befundung im Gesamten ( $p = 0,004$ ) und für Pulmologen ( $p = 0,0005$ ). Es gab keine signifikanten Unterschiede in Bezug auf die Patientendisposition oder die Therapie.

Zusammenfassend ließ sich feststellen, dass zuweisende Fachkliniker Strukturierte Befundberichte von CTPA-Untersuchungen im Vergleich zu Konventionellen Befundberichten als eindeutiger empfinden. Pulmologen schätzen Strukturierte Befundberichte aufgrund ihrer größeren Vollständigkeit und des höheren klinischen Nutzens. Die Strukturierte Befundung scheint das Management der Patienten mit LAE nicht zu verändern.

## 2.2 Strukturierte Befundung der CT-Angiographie der unteren Extremitäten

**Teilprojekt 2: Sabel BO, Plum JL, Czihal M, Lottspeich C, Schönleben F, Gäbel G, Schinner R, Schoeppe F, Meinel FG.** Structured Reporting of CT Angiography Runoff Examinations of the Lower Extremities. **Eur J Vasc Endovasc Surg.** doi: 10.1016/j.ejvs. 2018.01.026.

*Journal Impact Factor: 5,3*

Die Computertomographie-Angiographie (CTA) ist ein effektives, nicht-invasives Bildgebungsinstrument für die Beurteilung von peripheren arteriellen Erkrankungen und wird oft als Grundlage für weitere klinische Entscheidungen mit direkter Auswirkung auf die Patientenversorgung verwendet. Präzise Befundung und genaue Ergebnisinterpretation sind daher zwingend erforderlich. Das wachsende Interesse an IT-Lösungen zur Strukturierten Befundung und die qualitativ hochwertige Forschung im Feld der Strukturierten Befundung verdeutlichen den klinischen Bedarf an Strukturierten Befundungssystemen. In Teilprojekt 2 wurde die Hypothese aufgestellt, dass standardisierte Strukturierte Befundberichte von CT-Angiographie (CTA)-Untersuchungen bei Patienten mit chronischer oder akuter peripherer arterieller Verschlusskrankheit einen potenziellen Mehrwert im Vergleich zu frei diktierten Konventionellen Befundberichten haben.

In Teilprojekt 2 wurden für 52 Patienten, die eine CTA der unteren Extremitäten erhalten hatten, Konventionelle Befundberichte und Strukturierte Befundberichte erstellt. Die Stichprobengröße basierte auf Power-Kalkulationen mit einer Power von 95% und einem Signifikanzniveau von .007 (berichtigt um Multiples Testen). Die Konventionellen Befundberichte wurden in einer freien Textform diktiert; die Strukturierten Befundberichte enthielten eine konsistente Anordnung der Beobachtungen mit standardisierten Unterüberschriften. Die Konventionellen Befundberichte wurden mit den Strukturierten Befundberichten verglichen. Zwei Fachärzte für Gefäßmedizin und zwei Gefäßchirurgen bewerteten die Befundberichte hinsichtlich ihrer Zufriedenheit mit Eindeutigkeit, Vollständigkeit, klinischer Relevanz und Nützlichkeit sowie der Gesamtzufriedenheit. Zusätzlich trafen sie hypothetische Entscheidungen über weitere Untersuchungen und Therapien. Die Medianbewertungen wurden mit dem Wilcoxon Signed Rank Test und generalisierten linearen Mixed-Effects-Modellen verglichen.

Die Ergebnisse dieser Arbeit belegen, dass zuweisende Fachkliniker Strukturierte Befundberichte von CTA-Untersuchungen im Vergleich zu Konventionellen Befundberichten als eindeutiger (medianer Wert 9,0 vs. 7,0,  $p < .0001$ ), vollständiger (medianer Wert 9,0 vs. 7,5,  $p < .0001$ ), klinisch relevanter (medianer Wert 9,0 vs. 8,0,  $p < .0001$ ) und hilfreicher (medianer Wert 9,0 vs. 8,0,  $p < .0001$ ) wahrnehmen. Die Strukturierten Befundberichte erreichten im Vergleich zu den Konventionellen Befundberichten sowohl bei Fachärzten für Gefäßmedizin als auch Gefäßchirurgen eine höhere Gesamtzufriedenheit (medianes Rating 9,0 vs. 7,0,  $p < .0001$ ). Allerdings scheint die Strukturierte Befundung bei Patienten mit bekannter oder vermuteter peripherer arterieller Verschlusskrankheit oder akuter embolischer Ischämie der Extremitäten keinen Einfluss auf weiterführende Diagnostik oder die Therapie zu haben.

### 2.3 KI-Unterstützung zur Befundung von CT-Kontrollen thorakaler Aortenaneurysmen

**Teilprojekt 3:** Rueckel J, Reidler P, Fink N, Sperl J, Geyer T, Fabritius MP, Ricke J, Ingrisich M, Sabel BO. Artificial intelligence assistance improves reporting efficiency of thoracic aortic aneurysm CT follow-up. **Eur J Radiol.** doi: 10.1016/j.ejrad. 2020.109424.

*Journal Impact Factor: 2,7*

Thorakale Aortenaneurysmata (TAA) der aufsteigenden Aorta mit einem Durchmesser über 6 cm rupturieren oder dissezieren mit einer jährlichen Rate von 4 % (9). Expansionsraten von mehr als 5,5 mm pro Jahr wurden als starke Risikofaktoren für eine spontane TAA-Ruptur beschrieben (53). Gemäß den ESC-Leitlinien sollte bei TAAs größer als 5,5 cm ein Ersatz der aufsteigenden Aorta in Erwägung gezogen werden. Eine sorgfältige Verlaufskontrolle der Aneurysmamorphanie mittels kontrastverstärkter elektrokardiogramm-gesteuerter (EKG-gesteuerter) CT ist folglich von entscheidender Bedeutung (10). Dennoch ist vielfach eine erhebliche Fehleranfälligkeit der Aortendurchmessermessung beschrieben (4, 10, 26, 29, 31, 35): Die Vergleichbarkeit von Follow-Up Untersuchungen erfordert Messungen an exakt denselben anatomischen Landmarkenpositionen, wobei jede Durchmessermessung entscheidend von der korrekten Ausrichtung der multiplanaren Rekonstruktion orthogonal zur Aortenmittellinie abhängt. Darüber hinaus können Pulsations-Artefakte und die Elastizität der Aorta die Beurteilung selbst bei EKG-gesteuerten CTs erschweren (33). Eine hochpräzise Verlaufsbeurteilung von Aortenaneurysmen erweist sich daher als ein anspruchsvolles und zeitaufwändiges Verfahren mit hoher therapeutischer Relevanz.

In Teilprojekt 3 evaluierten wir den Prototyp eines KI-Algorithmus zur detaillierten Beurteilung der Aorta. Der Algorithmus quantifiziert automatisch den Aortendurchmesser an neun anatomischen Orientierungspunkten gemäß der Leitlinie der European Society of Cardiology von 2014 (10) und der Leitlinie der American Heart Association von 2010 (20). Wir stellen die Hypothese auf, dass eine KI-gestützte Befundung die Inter-Reader-Variabilität für die Vermessung der Aorta reduziert und dennoch zeiteffizienter ist.

Der Einfluss der AI-Assistenz auf die Effizienz und Genauigkeit der Aortenaneurysma-Befundung nach den AHA / ESC-Richtlinien wurde anhand von 324 AI-Messungen und 1944

radiologischen Messungen quantifiziert: 18 Aortenaneurysma-Patienten mit jeweils zwei CT-Scans (arterielle Kontrastphase, elektrokardiogrammgesteuert) mit einem Intervall von mindestens sechs Monaten wurden eingeschlossen. Ein Facharzt für Radiologie und zwei Assistenzärzte (8/4/2 Jahre Erfahrung in der vaskulären Bildgebung) beurteilten unabhängig voneinander den Aortendurchmesser an neun Landmarken. Die Aneurysmaausdehnung wurde mit den ursprünglichen CT-Befundberichten verglichen. Nach einer dreiwöchigen Washout-Periode wurden die CTs auf Basis der grafisch dargestellten AI-Messungen neu begutachtet.

Die KI-Assistenz reduzierte die Variabilität der Messungen des Aortendurchmessers um 42,5% (0,42 / 1,16 mm mit / ohne KI-Assistenz, Mittelwert aller Patienten und Landmarkenpositionen, signifikante Reduktion für 6 von 9 Mess-positionen). Die Erfassung der Aneurysmaprogressions zwischen den Untersuchungen variierte bei konventioneller und KI-gestützter Quantifizierung nur gering (Mittelwert von 0,75 mm über alle Patienten / Landmarkenpositionen); die KI-gestützten Messungen übertrafen nicht die Inter-Reader-Variabilität der Radiologen. Die KI-Unterstützung reduzierte die mittlere Befundungszeit um 63 % von 13:01 auf 04:46 min, einschließlich manueller Korrekturen der KI-Messungen, welche bei 33,6 % aller Messungen durchgeführt wurden. Die zeitaufwändigen leitlinienkonformen Aortenmessungen ergaben bei 80 % der Aneurysmen, die im ursprünglichen Befund als auf die Aorta ascendens beschränkt beschrieben worden waren, eine zusätzliche Beteiligung der Aortenwurzel oder des Aortenbogens.

Wir schlussfolgern in Teilprojekt 3, dass die detaillierte, leitlinienkonforme Aortenbeurteilung zeitaufwendig, aber entscheidend für die genaue Darstellung der Aneurysmaausdehnung ist. Wir haben gezeigt, dass die KI-Assistenz ein vielversprechendes Werkzeug zur Verbesserung der Zeiteffizienz der radiologischen Befundung ist und hierdurch die Etablierung einer leitlinienkonformen Aorten-beurteilung in der klinischen Routine unterstützen könnte. Da die KI-Assistenz die messbasierte Entscheidungsfindung des Radiologen erheblich beeinträchtigt, ist eine hohe Genauigkeit der KI-gestützten Messungen entscheidend und sollte grundsätzlich innerhalb der Inter-Reader-Variabilität von Radiologen liegen. Weitere Fortschritte wären mutmaßlich durch eine Algorithmus-basierte sektionale Aortenluminalvolumetrie, anstelle der bisherigen isolierten Durchmesser-Messung in einer einzigen orthograden Messebene, zu

erreichen. Hierdurch wäre mutmaßlich eine deutliche Verbesserung der diagnostischen Leistung möglich, welche mit hoher Wahrscheinlichkeit sogar die von durch Inter-Reader- (und wahrscheinlich auch Intra-Reader-) Variabilität stark eingeschränkten Genauigkeit von Radiologen übertrifft.

## 2.4 KI-unterstützte Befundung zur Reduzierung von übersehenen Pathologien in notfallmäßig durchgeführten Schockraum-CT Untersuchungen

**Teilprojekt 4:** Rueckel J, Sperl J, Kaestle S, Hoppe BF, Fink N, Rudolph J, Schwarze V, Strobl F, Ricke J, Ingrisch M, **Sabel BO**. Reduction of Missed Thoracic Findings in Emergency Whole-Body CT using Artificial Intelligence Assistance. **Quant Imaging Med Surg. 2021** (accepted)  
*Journal Impact Factor: 3,2*

Die radiologische Befundung von notfallmäßig durchgeführten Ganzkörper-CT-Scans ist zeitkritisch und birgt daher ein erhebliches Risiko, dass relevante Pathologien übersehen werden. Fortschritte auf dem Gebiet der künstlichen Intelligenz werfen die Frage auf, ob eine ergänzende halbautomatische Bildanalyse (KI-unterstützte Befundung) die Anzahl der übersehenen Nebenbefunde zeitsparend reduzieren könnte. Jüngste Studien haben gezeigt, dass KI generell in der Lage ist, Bildanalysen auf dem Niveau von Fachärzten durchzuführen (11, 16, 22, 28). Jedoch haben verfügbare KI-Algorithmen in der Regel einen sehr engen klinischen Fokus und werden oft nur hinsichtlich quantitativer diagnostischer Metriken mit der Leistung von Radiologen verglichen. Bisher konnten für den Einsatz in der Notfalldiagnostik vielversprechende Ergebnisse für die Erkennung von intrakraniellen Blutungen (13, 37), ischämischen Schlaganfällen (30) oder Wirbelkörperfrakturen (49) gezeigt werden.

In Teilprojekt 4 evaluierten wir klinisch den Prototypen einer Softwareplattform (einschließlich mehrerer noch nicht kommerziell erhältlicher KI-Algorithmen) zur Analyse von Thorax-CTs mit dem Fokus auf Lungenrundherde, Kardiomegalie, Koronarsklerose, Aortenaneurysmen und Wirbelkörperfrakturen. Unsere Hypothese war, dass eine bedeutende Anzahl der ursprünglich übersehenen Nebenbefunde in einer KI-unterstützten Befundung erkannt werden kann.

Diese retrospektive Proof-of-Concept-Studie umfasste 105 konsekutive Schockraum-Ganzkörper-CT-Scans. Die Bilddaten wurden von plattformgebundenen KI-Algorithmen analysiert und die Ergebnisse von Fachärzten der Radiologie überprüft und mit den originalen Radiologiebefunden verglichen. Wir identifizierten folgende ursprünglich übersehene Pathologien: 25 Patienten (23,8 %) mit Kardiomegalie oder grenzwertiger Herzgröße, 17 Patienten (16,2 %) mit koronaren Plaques, 34 Patienten (32,4 %) mit Aortenektasie, 2 Patienten

(1,9 %) mit Lungenläsionen, die als kontrollbedürftig eingestuft wurden, und 13 zunächst übersehene Wirbelkörperfrakturen (zwei davon mit akuter traumatischer Genese). Bedenklich war die hohe Anzahl falsch positiver oder nicht relevanter AI-basierter Befunde, insbesondere bei Lungenrundherden und Wirbelkörperfrakturen.

Unsere Ergebnisse zeigen das große Potenzial von KI-Konzepten zur Reduzierung von übersehenen Nebenbefunden in klinischen Notfallsituationen, in denen eine besonders zeitnahe radiologische Befundung erforderlich ist. Insbesondere die Integration verschiedener Algorithmen in einer gemeinsamen Softwarelösung bietet die Möglichkeit, durch die Kombination von spezialisierten und einzeln nur in einem engen Rahmen nutzbaren Algorithmen einen ganzheitlichen synergistischen Mehrwert für den klinischen Alltag zu generieren. Wenngleich Algorithmen zur Detektion der ursprünglich übersehenen Pathologien in erster Linie eine hohe Sensitivität erfordern, sind weitere Verbesserungen hinsichtlich der Spezifität notwendig, um auch die Anzahl der falsch-positiven oder nicht relevanten KI-Detektionen zu verringern, da diese andernfalls von Radiologen in einem zeitaufwändigen Verfahren aktiv bewertet und ausgeschlossen werden müssen.

## 2.5 KI-Algorithmus zur Erkennung von Pneumonien in Thorax-Röntgenbildern

**Teilprojekt 5:** Rueckel J, Kunz WG, Hoppe BF, Patzig M, Notohamiprodo M, Meinel FG, Cyran CC, Ingrisich M, Ricke J, **Sabel BO**. Artificial Intelligence Algorithm Detecting Lung Infection in Supine Chest Radiographs of Critically Ill Patients With a Diagnostic Accuracy Similar to Board-Certified Radiologists. **Crit Care Med**. doi: 10.1097/CCM.0000000000004397.

Journal Impact Factor: 7,4

Die Beurteilung von Lungenverdichtungen in Thorax-Liegendtaufnahmen bleibt sowohl für Radiologen als auch für Kliniker eine diagnostische Herausforderung. Die zugrundeliegenden Pathologien können unter anderem lokale Belüftungsstörungen (Atelektasen), Pleuraergüsse oder pneumonieverdächtige Konsolidierungen sein und begründen jeweils unterschiedliche Therapieentscheidungen. Auf künstlicher Intelligenz basierende Ansätze zur Beurteilung von Thorax-Liegendtaufnahmen könnten zu Zeitersparnis und größerer Befundeffizienz führen und möglicherweise auch als klinische Entscheidungshilfe für Nicht-Radiologen in

Krankenhäusern ohne radiologische 24/7-Versorgung dienen (2, 7). Voraussetzung für die klinische Anwendung ist die genaue quantitative Evaluation der Diagnosealgorithmen im Vergleich zu einem Referenzstandard. In Teilprojekt 5 wurde ein auf den öffentlichen NIH Chest-X-Ray14- und PLCO-Datensätzen trainierter KI-Algorithmus für die Erkennung von Pneumonie-verdächtigen Konsolidierungen, Atelektasen und Pleuraergüssen in Thorax-Liegendaufnahmen evaluiert. Die Leistung des KI-Algorithmus wurde mit der von radiologischen Fachärzten verglichen. Die Studie umfasste 166 Patienten, die innerhalb von 90 Minuten und ohne dazwischenliegende Intervention sowohl eine Thorax-Liegendaufnahme als auch eine CT-Untersuchung erhielten. Die CT-Untersuchungen dienten als Referenzstandard.

Die Zuverlässigkeit des Algorithmus wurde mit Fachärzten für Radiologie, die die Thorax-Röntgenbilder anhand von seitengetrennten Bewertungen für Pneumonien und Pleuraergüsse beurteilten (0 = nicht vorhanden, 1 = möglich und 2 = hochverdächtig), überprüft. Die Radiologen waren während der CT-Beurteilung für die Befunde der Thorax-Röntgenaufnahmen geblindet. Die Leistungen der Radiologen und des Algorithmus der künstlichen Intelligenz wurden durch eine Analyse der Receiver-Operating Characteristic Curve quantifiziert. Diagnostische Metriken (Sensitivität, Spezifität, positiver prädiktiver Wert, negativer prädiktiver Wert und Genauigkeit) wurden auf der Basis verschiedener Receiver-Operating-Characteristics-Punkte (ROC) berechnet. Die radiologische Beurteilung erzielte im Wesentlichen nicht-signifikant höhere Ergebnisse im Vergleich zum KI-Algorithmus: Area-under-the-Curve der künstlichen Intelligenz von 0,737 (0,659-0,815) versus Area-under-the-Curve der Radiologen von 0,779 (0,723-0,836), die diagnostischen Metriken der ROC unterschieden sich nicht signifikant. Bei der Erkennung von Pleuraergüssen gab es keinen signifikanten Leistungsunterschied zwischen den Radiologen und dem KI-Algorithmus: Area-under-the-Curve der KI von 0,740 (0,662-0,817) gegenüber der Area-under-the-Curve der Radiologen von 0,698 (0,646-0,749) mit ähnlichen diagnostischen Metriken für die ROC-Punkte. Bei der Erkennung von Pneumonien erzielte die radiologische Beurteilung also im Wesentlichen ähnliche Ergebnisse wie der AI-Algorithmus. Bei der Erkennung von Pleuraergüssen erreichten Radiologen und der KI-Algorithmus vergleichbare diagnostische Genauigkeiten.

Zusammenfassend halten wir KI für einen vielversprechenden Ansatz, um die hohe Anzahl von Thorax-Liegenaufnahmen in der klinischen Routine zu bewältigen oder in Krankenhäusern ohne radiologische Rund-um-die-Uhr Versorgung als klinische Entscheidungshilfe zur Reduktion von übersehenen Pathologien beizutragen. Darüber hinaus könnte eine KI-gestützte Triage durch Kennzeichnung vorselektierter Fälle eine priorisierte radiologische Befundung zeitkritischer Fälle ermöglichen und hierdurch Verzögerungen bei diesen vermeiden. Aufgrund des Risikos der Unterdiagnostik sollten KI-Diagnosealgorithmen jedoch nicht als alleinige Lösung verwendet werden.

## 2.6 Bedeutung von Thoraxdrainagen und Ausmaß der Pleuradehiszenz für die Pneumothorax-Erkennung im Röntgenthorax durch KI-Algorithmen

**Teilprojekt 6:** Rueckel J, Trappmann L, Schachtner B, Wesp P, Hoppe BF, Fink N, Ricke J, Dinkel J, Ingrisch M, **Sabel BO**. Impact of Confounding Thoracic Tubes and Pleural Dehiscence Extent on Artificial Intelligence Pneumothorax Detection in Chest Radiographs. *Invest Radiol*. doi: 10.1097/RLI.0000000000000707.

*Journal Impact Factor: 5,2*

Es wurde bereits mehrfach gezeigt, dass einige auf öffentlich verfügbaren Röntgenthorax-Datensätzen trainierte KI-Algorithmen die diagnostische Leistung von radiologischen Fachärzten, zum Beispiel hinsichtlich der Erkennung von Pneumonien oder Atelektasen, teilweise übertreffen (36, 40). In ähnlicher Weise haben mehrere, ebenfalls auf öffentlich verfügbaren Röntgenthorax-Datensätzen trainierte, KI-Algorithmen zur automatischen Pneumothorax-Erkennung in Röntgenbildern hohe diagnostische Genauigkeiten mit Areas under the Receiver-Operating-Characteristics (AUROCs) von bis zu 0,889 erzielt. Leider sind häufig genauere Charakteristika der verwendeten KI-Algorithmen, wie z.B. der Einfluss der Größe der Pneumothoraces oder potentiell die Leistung der Algorithmen beeinflussende Störfaktoren, nicht näher beschrieben. Insbesondere der Einfluss von einliegenden Thoraxdrainagen wurde diskutiert und bereits teilweise in Studien für das Algorithmen-Training ausgeschlossen (32, 48). In Teilprojekt 6 untersuchten wir, ob die Leistung von KI-Algorithmen von der Pneumothorax-Größe oder Bildmerkmalen, wie z.B. einliegenden Thoraxdrainagen, beeinflusst wird, wenn der Algorithmus ausschließlich anhand von auf klinischen Befunden beruhenden und unscharf gelabelten öffentlichen Datensätze trainiert wurde. Hierzu etablierten wir eine große radiologisch annotierte Benchmarking-Kohorte von Röntgenthorax-Aufnahmen, die eine detaillierte Subgruppenanalyse ermöglicht.

Wir identifizierten retrospektiv 6434 Röntgenthorax-Aufnahmen in Rückenlage, darunter 1652 Pneumothorax-positive Fälle und 4782 Pneumothorax-negative Fälle. Die Röntgenthorax-Aufnahmen wurden radiologisch hinsichtlich der Pneumothorax-Größe, der Pneumothorax-Lokalisation und der eingebrachten Thoraxdrainagen annotiert. Die diagnostische Leistung von zwei auf öffentlich verfügbaren Datensätzen mit Labels aus einer automatisierten Befundung trainierten KI-Algorithmen ("AI\_CheXNet" [Rajpurkar et al],

"AI\_1.5" [Guendel et al]) wurde quantifiziert. Die Trennschärfe der Algorithmen für die Pneumothorax-Detektion wurde durch die Fläche unter der Receiver Operating Characteristics (AUROC) quantifiziert, und die Signifikanzanalyse basierte auf dem entsprechenden 95%-Konfidenzintervall. Es wurde eine detaillierte Untergruppenanalyse durchgeführt, um den Einfluss der Pneumothorax-Größe und die durch eingefügte Thoraxdrainagen verursachten Störungseffekte zu quantifizieren.

Die Leistung des Algorithmus wurde wie folgt quantifiziert: Gesamtleistung mit AUROCs von 0,704 (AI\_1.5) / 0,765 (AI\_CheXNet) für unilaterale Pneumothoraces, AUROCs von 0. 666 (AI\_1,5) / 0,722 (AI\_CheXNet) für unilaterale Pneumothoraces kleiner als 1 cm und AUROCs von 0,735 (AI\_1,5) / 0,818 (AI\_CheXNet) für unilaterale Pneumothoraces größer als 2 cm. Die Subgruppenanalyse identifizierte Thoraxdrainagen als starke Confounder, die die Algorithmusleistung signifikant beeinflussen: Die diskriminative Leistung wird vollständig eliminiert, wenn Pneumothorax-positive Fälle ohne Thoraxdrainagen im Vergleich zu kontrollierten Pneumothorax-negativen Fällen mit eingelegten Thoraxdrainagen analysiert werden. Im Gegensatz dazu stiegen die AUROCs auf bis zu 0,875 (AI\_CheXNet) für große Pneumothorax-positive Fälle mit eingefügten Thoraxdrainagen in Bezug auf Kontrollfälle ohne Thoraxdrainagen.

Zusammenfassend haben wir am Beispiel der KI-basierten Pneumothorax-Detektion die dringende Notwendigkeit aufgezeigt, dass Trainingsdaten von KI-Algorithmen sorgfältig auf möglicherweise störende Bildmerkmale zu kontrollieren sind, da diese andernfalls in einer detaillierten Subgruppenanalyse die Algorithmusleistung vollständig eliminieren könnten. Wenn störende Bildmerkmale während des Trainings schwer zu vermeiden sind (z.B. ein relativ hoher Anteil von Thoraxdrainagen in Pneumothorax -positiven Röntgenbildern), müssen sie unbedingt beim Algorithmus-Training berücksichtigt werden, z. B. durch explizite Annotationen. Klinisch aussagekräftige und radiologisch annotierte Benchmarking-Kohorten, wie sie in dieser Arbeit etabliert und beschrieben wurden, sind die Voraussetzung für eine effektive und kontinuierliche Algorithmenentwicklung, aber auch zur Charakterisierung bereits zugelassener Algorithmen, deren Eigenschaften zum Teil unzureichend beschrieben sind.



## 2.7 Prognostischer Wert des CT-Thorax für die invasive Beatmungstherapie bei COVID-19-Pneumonien

**Teilprojekt 7:** Gresser E, Rueckel J, Pühr-Westerheide D, Schwarze V, Fink N, Kunz WG, Wassilowsky D, Irlbeck M, Ricke J, Ingrisich M, **Sabel BO**. Prognostic Value of Admission Chest CT Findings for Invasive Ventilation Therapy in COVID-19 Pneumonia. **Diagnostics** doi: 10.3390/diagnostics10121108.

Journal Impact Factor: 3,1

Meta-Analysen ergaben, dass etwa 15-30% der hospitalisierten COVID-19 Patienten einen kritischen Krankheitsverlauf mit dem Erfordernis einer Behandlung auf der Intensivstation (ICU) und invasiver Beatmung ausbilden. Das Ziel von Teilprojekt 7 war es, die in unserem Krankenhaus hospitalisierte Patientenkohorte von bestätigten COVID-19 Fällen während der ersten Welle der Pandemie in Deutschland (Anfang März bis Anfang April 2020) zu analysieren und Risikofaktoren für die Invasive-Ventilations-Therapie (IV-Therapie) auf der Intensivstation zu ergründen. Wir stellten die Hypothese auf, dass prognostische Informationen über den Krankheitsverlauf aus Thorax-CT-Untersuchungen bei Aufnahme des Patienten gewonnen werden können.

Es wurde eine retrospektive Single-Center-Studie durchgeführt, die alle während der ersten drei Monate der Pandemie an unserem Krankenhaus aufgenommenen Patienten mit PCR-bestätigter COVID-19-Erkrankung und Thorax-CT-Untersuchung bei Krankenhausaufnahme einschloss (n = 69). Anhand der klinischen Patientendaten und der CT-Bildgebungsmerkmale wurde ein ordinaler 10-Punkte-Risikoscore entwickelt und dessen diagnostisches Potenzial zur Unterscheidung eines schweren (IV-Gruppe) von einem moderateren Verlauf (Nicht-IV-Gruppe) der Erkrankung überprüft.

Als Ergebnis stellten wir ein ordinales 10-Punkte-Scoring-System für die IV-Risikostratifizierung vor, das sowohl aus demographischen Merkmalen als auch aus Merkmalen der Bildgebung besteht. Es wurden alle bildgebenden Merkmale einbezogen, die signifikant häufiger in der IV-Kohorte auftraten oder ein Odds Ratio von >2,0 für IV aufwiesen. Um eine

Übergewichtung seltener Befunde im Scoring-System zu vermeiden wurden nur Merkmale mit einer Prävalenz von mehr als 10% innerhalb der Kohorte aufgenommen. Die einbezogenen Verteilungsmerkmale spiegeln die Lungenbeteiligung wider, welche sich in früheren Studien als prognostisch entscheidender Faktor erwiesen hat (23, 51, 54). Häufige Befunde in der Bildgebung der COVID-19-Pneumonie in beiden Gruppen waren Milchglastrübungen (91,3%), Konsolidierungen (53,6%) und Crazy-Paving-Muster (31,9%). Merkmale späterer Stadien wie etwa subpleurale Bänder wurden signifikant häufiger in der IV-Gruppe beobachtet (52,2% versus 26,1%,  $p = 0,032$ ). Neben für fortgeschrittene Stadien der COVID-Erkrankung sprechenden Merkmalen wurden auch solche aufgenommen, die für kardiovaskuläre Komorbiditäten oder systemische Entzündungen sprechen. Aufgrund seiner signifikanten Bedeutung für IV innerhalb unserer Kohorte wurde das männliche Geschlecht als demografischer Parameter einbezogen. Ein hohes Lebensalter wurde als wichtiger Risikofaktor für den Schweregrad der Erkrankung beschrieben und deshalb in den Score aufgenommen (24). Unser Scoring-System erwies sich als geeignet, mit einer AUC von 0,89 [95% CI 0,81-0,96] eindeutig zwischen Patienten mit hohem Risiko für eine invasive Beatmung (IV) und Patienten mit moderaten Krankheitsverläufen (non-IV) zu unterscheiden. Es stellte sich heraus, dass ein Cut-off-Wert von 5 Punkten die Sensitivität und Spezifität maximiert und damit die knappen Kapazitäten der Intensivstation während dieser Krise am zuverlässigsten berücksichtigt (Sensitivität von 0,81 und Spezifität von 0,85). Ein Cut-Off von 4 Punkten könnte jedoch für eine Minimierung falsch negativ bewerteter Patienten sorgen, die eine Infusion benötigen und daher frühzeitig enger überwacht und möglicherweise anders behandelt werden sollten (Sensitivität  $> 0,9$ , Spezifität  $> 0,65$ ). Unter Verwendung von Informationen, die bei der Befundung durch den Radiologen direkt zugänglich sind, erwies sich ein einfacher Risikoscore als zuverlässig zur Unterscheidung zwischen IV- und Nicht-IV-Gruppe (AUC: 0,89 [95% CI 0,81-0,96],  $p < 0,001$ ).

Der hier vorgestellte Score kann durch seinen Einsatz in der radiologischen Routinebefundung als „imaging biomarker“ einen wertvollen Beitrag zur Gesamtbeurteilung des Patientenrisikos leisten und die effektive Patienten-Disposition und das Ressourcenmanagement in Krankenhäusern und Krankenhausnetzwerken unterstützen.



## 2.8 Differenzierung zwischen COVID-19, Influenza und anderen Pneumonien mittels CT

**Teilprojekt 8:** Rueckel J, Fink N, Kaestle S, Stueber AT, Schwarze V, Gresser EK, Hoppe BF, Rudolph J, Kunz WG, Ricke J, **Sabel BO** COVID-19 Pandemic and Upcoming Influenza Season—Does an Expert's Computed Tomography Assessment Differentially Identify COVID-19, Influenza and Pneumonias of Other Origin? **J. Clin. Med.** 2021, 10(1), 84; doi.org/10.3390/jcm10010084

*Journal Impact Factor: 3,3*

Mehrere Studien haben gezeigt, dass die Thorax-CT zum Nachweis von COVID-19, insbesondere in frühen Erkrankungsstadien, im Vergleich zur RT-PCR höhere Sensitivitäten aufweist (1, 12). Die meisten dieser Studien lieferten jedoch keine ausgewogenen und ausreichend großen Untergruppen, um die Spezifität von COVID-19-bezogenen CT-Bildgebungsmerkmalen zu quantifizieren oder gingen nicht spezifisch auf andere relevante Erreger, z. B. ausschließlich mit einer Influenza-Infektion assoziierte atypische Pneumonien, ein (3, 25, 50). Im Gegensatz zu diesen Studien beschreiben wir nicht bereits bekannte erregerbezogene CT-Bildmuster, sondern quantifizieren die Leistung des Radiologen zur differenzierten erregerspezifischen Verdachtsauslösung mit dem Fokus auf Influenza A/B und SARS-CoV-2. Dieser Ansatz zielt darauf ab, den diagnostischen Mehrwert einer schnell verfügbaren Thorax-CT für die Pneumonie-Diagnostik durch erfahrene Radiologen, vor dem Hintergrund der kommende Grippesaison und zugleich währenden COVID-19-Pandemie, zu bewerten.

Insgesamt wurden 201 Pneumonie-CTs identifiziert und anhand der RT-PCR in Untergruppen eingeteilt: 78 COVID-19-CTs, 65 Influenza-CTs und 62 Nicht-COVID-19-Nicht-Influenza (NCNI)-CTs. Drei Radiologen (geblindet gegenüber den RT-PCR-Ergebnissen) erhoben einen erregerspezifischen Verdacht (getrennt für COVID-19, Influenza, bakterielle Pneumonie und Pilzpneumonie) gemäß den folgenden Befundungs-Scores: 0-nicht typisch/1-möglich/2-hochgradig verdächtig. Die diagnostischen Leistungen wurden mit der RT-PCR als Referenzstandard ermittelt. Die Abhängigkeiten der Erregerverdachts-Scores der Radiologen wurden mit dem Pearson's Chi<sup>2</sup>-Test für Unabhängigkeit charakterisiert.

Je nachdem, ob der intermediäre Befundungs-Score 1 als positiv oder negativ gewertet wurde, klassifizierten Radiologen 83%-85% (vs. NCNI)/79%-82% (vs. Influenza) der COVID-19-Fälle korrekt (Sensitivität bis zu 94%). Im Gegensatz dazu klassifizierten die Radiologen nur 52%-56% (vs. NCNI)/50-60% (vs. COVID-19) der Influenza-Fälle richtig. Das COVID-19-Scoring war spezifischer als das Influenza-Scoring, verglichen mit dem Verdacht auf eine bakterielle Infektion oder Pilzinfektion.

Zusammenfassend bestätigten wir die diagnostische Aussagekraft der Thorax-CT zur sensitiven Identifizierung einer COVID-19-Pneumonie mit dem Fokus auf frühe Infektionsstadien, auch im Vergleich mit anderen atypischen Pneumonien wie der Influenza-Infektion. Während der kommenden Influenza-Saison könnte die Thorax-CT helfen, SARS-CoV-2-Infektion verdächtige Pneumonien schnell zu identifizieren und so die häufig unzuverlässige und zeitaufwendige SARS-CoV-2 RT-PCR zu unterstützen. Im Gegensatz zu SARS-CoV-2 kann eine Influenza-Pneumonie aufgrund unspezifischer Bildmerkmale nicht zuverlässig mittels Thorax-CT identifiziert werden. Diese diagnostische Lücke kann jedoch leicht durch PCR-Tests geschlossen werden, da diese für den Influenza-Nachweis deutlich zuverlässiger sind als für den Nachweis von SARS-CoV-2.

### 3 Literaturverzeichnis

1. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020;296(2):E32-e40.
2. Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated Triage of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology*. 2019;291(1):196-202.
3. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. *Radiology*. 2020;296(2):E46-e54.
4. Banno H, Kobeiter H, Brossier J, Marzelle J, Presles E, Becquemin JP. Inter-observer variability in sizing fenestrated and/or branched aortic stent-grafts. *Eur J Vasc Endovasc Surg*. 2014;47(1):45-52.
5. Brook OR, Brook A, Vollmer CM, Kent TS, Sanchez N, Pedrosa I. Structured reporting of multiphasic CT for pancreatic cancer: potential effect on staging and surgical planning. *Radiology*. 2015;274(2):464-72.
6. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, et al. The ACR BI-RADS experience: learning from history. *J Am Coll Radiol*. 2009;6(12):851-60.
7. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *Eur J Radiol*. 2020;123:108774.
8. Collard MD, Tellier J, Chowdhury AS, Lowe LH. Improvement in reporting skills of radiology residents with a structured reporting curriculum. *Academic radiology*. 2014;21(1):126-33.
9. Elefteriades JA. Indications for aortic replacement. *J Thorac Cardiovasc Surg*. 2010;140(6 Suppl):S5-9; discussion S45-51.
10. Erbel R, Aboyans V, Boileau C, Bossone E, Bartolomeo RD, Eggebrecht H, et al. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: Document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). *Eur Heart J*. 2014;35(41):2873-926.
11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-8.
12. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology*. 2020;296(2):E115-e7.
13. Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology*. 2020;62(3):335-40.
14. Gresser E, Rueckel J, Pühr-Westerheide D, Schwarze V, Fink N, Kunz WG, et al. Prognostic Value of Admission Chest CT Findings for Invasive Ventilation Therapy in COVID-19 Pneumonia. *Diagnostics (Basel)*. 2020;10(12).
15. Grieve FM, Plumb AA, Khan SH. Radiology reporting: a general practitioner's perspective. *Br J Radiol*. 2010;83(985):17-22.
16. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama*. 2016;316(22):2402-10.
17. Hager A, Kaemmerer H, Rapp-Bernhardt U, Blücher S, Rapp K, Bernhardt TM, et al. Diameters of the thoracic aorta throughout life as measured with helical computed tomography. *J Thorac Cardiovasc Surg*. 2002;123(6):1060-6.
18. Hannuksela M, Lundqvist S, Carlberg B. Thoracic aorta--dilated or not? *Scand Cardiovasc J*. 2006;40(3):175-8.
19. Hawkins CM, Hall S, Zhang B, Towbin AJ. Creation and implementation of department-wide structured reports: an analysis of the impact on error rate in radiology reports. *J Digit Imaging*. 2014;27(5):581-7.
20. Hiratzka LF, Bakris GL, Beckman JA, Bersin RM, Carr VF, Casey DE, Jr., et al. 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with Thoracic Aortic Disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, American College of

- Radiology, American Stroke Association, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society of Interventional Radiology, Society of Thoracic Surgeons, and Society for Vascular Medicine. *Circulation*. 2010;121(13):e266-369.
21. Johnston KW, Rutherford RB, Tilson MD, Shah DM, Hollier L, Stanley JC. Suggested standards for reporting on arterial aneurysms. Subcommittee on Reporting Standards for Arterial Aneurysms, Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery and North American Chapter, International Society for Cardiovascular Surgery. *J Vasc Surg*. 1991;13(3):452-8.
  22. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*. 2017;284(2):574-82.
  23. Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur Radiol*. 2020;30(8):4407-16.
  24. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *J Allergy Clin Immunol*. 2020;146(1):110-8.
  25. Li Y, Xia L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. *AJR Am J Roentgenol*. 2020;214(6):1280-6.
  26. Lu TL, Rizzo E, Marques-Vidal PM, Segesser LK, Dehmeshki J, Qanadli SD. Variability of ascending aorta diameter measurements as assessed with electrocardiography-gated multidetector computerized tomography and computer assisted diagnosis software. *Interact Cardiovasc Thorac Surg*. 2010;10(2):217-21.
  27. Marcovici PA, Taylor GA. Journal Club: Structured radiology reports are more complete and more effective than unstructured reports. *AJR Am J Roentgenol*. 2014;203(6):1265-71.
  28. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafiyan H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94.
  29. Mora CE, Marcus CD, Barbe CM, Ecarnot FB, Long AL. Maximum Diameter of Native Abdominal Aortic Aneurysm Measured by Angio-Computed Tomography: Reproducibility and Lack of Consensus Impacts on Clinical Decisions. *Aorta (Stamford)*. 2015;3(2):47-55.
  30. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg*. 2020;12(2):156-64.
  31. Oshin OA, England A, McWilliams RG, Brennan JA, Fisher RK, Vallabhaneni SR. Intra- and interobserver variability of target vessel measurement for fenestrated endovascular aneurysm repair. *J Endovasc Ther*. 2010;17(3):402-7.
  32. Park S, Lee SM, Kim N, Choe J, Cho Y, Do KH, et al. Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. *Eur Radiol*. 2019;29(10):5341-8.
  33. Parodi J, Berguer R, Carrascosa P, Khanafer K, Capunay C, Wizauer E. Sources of error in the measurement of aortic diameter in computed tomography scans. *J Vasc Surg*. 2014;59(1):74-9.
  34. Plumb AA, Grieve FM, Khan SH. Survey of hospital clinicians' preferences regarding the format of radiology reports. *Clin Radiol*. 2009;64(4):386-94; 95-6.
  35. Quint LE, Liu PS, Booher AM, Watcharotone K, Myles JD. Proximal thoracic aortic diameter measurements at CT: repeatability and reproducibility according to measurement method. *Int J Cardiovasc Imaging*. 2013;29(2):479-88.
  36. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018;15(11):e1002686.
  37. Rao B, Zohrabian V, Cedeno P, Saha A, Pahade J, Davis MA. Utility of Artificial Intelligence Tool as a Prospective Radiology Peer Reviewer - Detection of Unreported Intracranial Hemorrhage. *Academic radiology*. 2021;28(1):85-93.
  38. Roman MJ, Devereux RB, Kramer-Fox R, O'Loughlin J. Two-dimensional echocardiographic aortic root dimensions in normal children and adults. *Am J Cardiol*. 1989;64(8):507-12.

39. Rueckel J, Fink N, Kaestle S, Stüber T, Schwarze V, Gresser E, et al. COVID-19 Pandemic and Upcoming Influenza Season-Does an Expert's Computed Tomography Assessment Differentially Identify COVID-19, Influenza and Pneumonias of Other Origin? *J Clin Med.* 2020;10(1).
40. Rueckel J, Kunz WG, Hoppe BF, Patzig M, Notohamiprodjo M, Meinel FG, et al. Artificial Intelligence Algorithm Detecting Lung Infection in Supine Chest Radiographs of Critically Ill Patients With a Diagnostic Accuracy Similar to Board-Certified Radiologists. *Crit Care Med.* 2020.
41. Rueckel J, Reidler P, Fink N, Sperl J, Geyer T, Fabritius MP, et al. Artificial intelligence assistance improves reporting efficiency of thoracic aortic aneurysm CT follow-up. *Eur J Radiol.* 2021;134:109424.
42. Rueckel J SJ, Kaestle S, Hoppe BF, Fink N, Rudolph J, Schwarze V, Strobl F, Ricke J, Ingrisich M, Sabel BO. Reduction of Missed Thoracic Findings in Emergency Whole-Body CT using Artificial Intelligence Assistance. *Quant Imaging Med Surg* 2021 (accepted). 2021.
43. Rueckel J, Trappmann L, Schachtner B, Wesp P, Hoppe BF, Fink N, et al. Impact of Confounding Thoracic Tubes and Pleural Dehiscence Extent on Artificial Intelligence Pneumothorax Detection in Chest Radiographs. *Invest Radiol.* 2020;55(12):792-8.
44. Sabel BO, Plum JL, Czihal M, Lottspeich C, Schönleben F, Gäbel G, et al. Structured Reporting of CT Angiography Runoff Examinations of the Lower Extremities. *Eur J Vasc Endovasc Surg.* 2018;55(5):679-87.
45. Sabel BO, Plum JL, Kneidinger N, Leuschner G, Koletzko L, Raziorrouh B, et al. Structured reporting of CT examinations in acute pulmonary embolism. *J Cardiovasc Comput Tomogr.* 2017;11(3):188-95.
46. Schwartz LH, Panicek DM, Berk AR, Li Y, Hricak H. Improving communication of diagnostic radiology findings through structured reporting. *Radiology.* 2011;260(1):174-81.
47. Sistrom CL, Langlotz CP. A framework for improving radiology reporting. *J Am Coll Radiol.* 2005;2(2):159-67.
48. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. *PLoS Med.* 2018;15(11):e1002697.
49. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med.* 2018;98:8-15.
50. Wang H, Wei R, Rao G, Zhu J, Song B. Characteristic CT findings distinguishing 2019 novel coronavirus disease (COVID-19) from influenza pneumonia. *Eur Radiol.* 2020;30(9):4910-7.
51. Wang Y, Dong C, Hu Y, Li C, Ren Q, Zhang X, et al. Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. *Radiology.* 2020;296(2):E55-e64.
52. Weiss DL, Langlotz CP. Structured reporting: patient care enhancement or productivity nightmare? *Radiology.* 2008;249(3):739-47.
53. Yiu RS, Cheng SW. Natural history and risk factors for rupture of thoracic aortic arch aneurysms. *J Vasc Surg.* 2016;63(5):1189-94.
54. Zhou Z, Guo D, Li C, Fang Z, Chen L, Yang R, et al. Coronavirus disease 2019: initial chest CT findings. *Eur Radiol.* 2020;30(8):4398-406.