

Grenzen künstlicher Intelligenz in der Notfallbefundung – eine Leistungsanalyse eines kommerziellen, computerbasierten Algorithmus zur Detektion von Lungenarterienembolien

Present Limitations of Artificial Intelligence in the Emergency Setting – Performance Study of a Commercial, Computer-Aided Detection Algorithm for Pulmonary Embolism

Autoren

Katharina Müller-Peltzer¹, Lena Kretzschmar², Giovanna Negrão de Figueiredo², Alexander Crispin³, Robert Stahl⁴, Fabian Bamberg¹, Christoph Gregor Trumm⁴

Institute

- 1 Klinik für Diagnostische und Interventionelle Radiologie, Albert-Ludwigs-Universität Freiburg Medizinische Fakultät, Freiburg, Deutschland
- 2 Klinik und Poliklinik für Radiologie, Ludwig-Maximilians-Universität, München, Deutschland
- 3 Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Klinikum der Universität München-Großhadern, München, Deutschland
- 4 Institut für Diagnostische und Interventionelle Neuroradiologie, Klinikum der Universität München-Großhadern, München, Deutschland

Key words

pulmonary embolism, CAD, CT pulmonary angiogram, computer-aided detection

eingereicht 31.12.2020

akzeptiert 04.05.2021

online publiziert 05.08.2021

Bibliografie

Fortschr Röntgenstr 2021; 193: 1436–1444

DOI 10.1055/a-1515-2923

ISSN 1438-9029

© 2021. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Korrespondenzadresse

Dr. Katharina Müller-Peltzer

Klinik für Diagnostische und Interventionelle Radiologie, Albert-Ludwigs-Universität Freiburg, Hugstetter Str. 55, 79106 Freiburg im Breisgau, Germany

Tel.: +49/7 61/27 03 81 90

Katharina.Mueller-Peltzer@uniklinik-freiburg.de

ZUSAMMENFASSUNG

Ziel Da zunehmend Anwendungen künstlicher Intelligenz im klinischen Alltag implementiert werden, war das Ziel unserer Studie, die Leistung eines kommerziellen, computerassistierten Detektionsalgorithmus für Lungenarterienembolien in Computertomografie-Pulmonalisangiografien im Rahmen einer Notfallbefundung zu analysieren.

Material und Methoden In diese retrospektive Studie wurden alle Computertomografie-Pulmonalisangiografien eingeschlossen, die über einen Zeitraum von 36 Monaten in einer großen deutschen Notaufnahme gefahren wurden. Alle eingeschlossenen Untersuchungen wurden von 2 in Notfallradiologie erfahrenen Radiologen bewertet, um einen Referenzstandard festzulegen. Die Originalbefunde wurden mit den Ergebnissen des computerassistierten Detektionsalgorithmus in Bezug auf die Erkennung lobärer, segmentaler und subsegmentaler Lungenarterienembolien verglichen. Für alle Ergebnisse des computerbasierten Detektionsalgorithmus wurde die zugrunde liegende Pathologie definiert. Falsch positive Ergebnisse des Detektionsalgorithmus wurden mit dem Kontrast-Rausch-Verhältnis verglichen.

Ergebnisse Im Rahmen der Referenzbefundung wurden 504 Emboli in 182 von insgesamt 1229 Patienten (49 % männlich, 10–97 Jahre) detektiert. Der computerassistierte Detektionsalgorithmus zeigte insgesamt 3331 Ergebnisse an, hiervon waren 258 (8 %) richtig positiv und 3073 (92 %) falsch positiv. Die Sensitivität des Detektionsalgorithmus betrug 47 % (95 %-Konfidenzintervall 33–61 %) für lobäre und 50 % (95 %-Konfidenzintervall 43–56 %) für subsegmentale Lungenarterienembolien. Im Durchschnitt wurden 2,25 Ergebnisse pro Untersuchung (Median 2, 0–25) angezeigt. Es gab keine signifikante Korrelation zwischen der Anzahl der falsch positiven Ergebnisse und dem Kontrast-Rausch-Verhältnis (Spearman's Rangkorrelationskoeffizient = 0,09). Weichteilgewebe (61,0 %) und Pulmonalvenen (24,1 %) waren die häufigsten Ursachen für falsch positive Ergebnisse.

Schlussfolgerung Die Anwendung des getesteten computerassistierten Detektionsalgorithmus auf ein Patientenkollektiv einer großen Notaufnahme deckt die Schwachstellen des

Algorithmus auf, die im Rahmen zukünftiger Projektentwicklungen optimiert werden sollten.

Kernaussagen:

- Die computertomografische Pulmonalisangiografie ist eine häufige Anforderung in der Notfallversorgung.
- Computerassistierte Detektionsalgorithmen können bei der Bildanalyse helfen.
- Die Detektionsalgorithmen weisen noch Schwächen bei falsch positiven und falsch negativen Befunden auf.
- Radiolog*innen müssen sich der Limitationen der Detektionsalgorithmen bewusst sein.

Zitierweise

- Müller-Peltzer K, Kretschmar L, Negrão de Figueiredo G et al. Present Limitations of Artificial Intelligence in the Emergency Setting – Performance Study of a Commercial, Computer-Aided Detection Algorithm for Pulmonary Embolism. *Fortschr Röntgenstr* 2021; 193: 1436–1444

ABSTRACT

Purpose Since artificial intelligence is transitioning from an experimental stage to clinical implementation, the aim of our study was to evaluate the performance of a commercial, computer-aided detection algorithm of computed tomography pulmonary angiograms regarding the presence of pulmonary embolism in the emergency room.

Materials and Methods This retrospective study includes all pulmonary computed tomography angiogram studies

performed in a large emergency department over a period of 36 months that were analyzed by two radiologists experienced in emergency radiology to set a reference standard. Original reports and computer-aided detection results were compared regarding the detection of lobar, segmental, and subsegmental pulmonary embolism. All computer-aided detection findings were analyzed concerning the underlying pathology. False-positive findings were correlated to the contrast-to-noise ratio.

Results Expert reading revealed pulmonary embolism in 182 of 1229 patients (49 % men, 10–97 years) with a total of 504 emboli. The computer-aided detection algorithm reported 3331 findings, including 258 (8 %) true-positive findings and 3073 (92 %) false-positive findings. Computer-aided detection analysis showed a sensitivity of 47 % (95 %CI: 33–61 %) on the lobar level and 50 % (95 %CI 43–56 %) on the subsegmental level. On average, there were 2.25 false-positive findings per study (median 2, range 0–25). There was no significant correlation between the number of false-positive findings and the contrast-to-noise ratio (Spearman’s Rank Correlation Coefficient = 0.09). Soft tissue (61.0 %) and pulmonary veins (24.1 %) were the most common underlying reasons for false-positive findings.

Conclusion Applied to a population at a large emergency room, the tested commercial computer-aided detection algorithm faced relevant performance challenges that need to be addressed in future development projects.

Einleitung

Die akute Lungenarterienembolie (LAE) ist eine häufige Erkrankung mit potenziell hoher Morbidität und Mortalität [1]. In internationalen, epidemiologischen Studien wurde eine jährliche Inzidenz der akuten LAE von 39–115/100 000 Personen ermittelt [2]. In den letzten Jahren wurde zudem eine steigende Inzidenzrate akuter Lungenarterienembolien verzeichnet. Dies ist insbesondere auf einen vermehrten Einsatz der Computertomografie (CT) in der Notfalldiagnostik zurückzuführen, die mittlerweile auch den Nachweis kleiner, weit peripher gelegener Lungenarterienembolien auf Subsegmentebene ermöglicht. Trotz Scoring-Systemen zur Abschätzung der Wahrscheinlichkeit einer LAE werden somit auch vermehrt Patient*innen untersucht, deren Symptomatik auf anderen Ursachen basiert [3–5]. Mit jeder CT-Pulmonalisangiografie (CTPA) wird eine große Menge an Bilddaten generiert. Insbesondere die genaue Evaluation der distalen Subsegmente stellt in Zusammenschau mit den wachsenden CT-Untersuchungszahlen eine Herausforderung in der Notfallbefundung dar. Automatisierte Detektionsalgorithmen (computer-aided detection, CAD) sollen bei derartig zeitaufwendigen Bildanalysen unterstützen können. Während CAD-Systeme für die Lungenrundherddetektion bereits gut etabliert sind, finden LAE-CAD noch keine Anwendung in der klinischen Routine. Dies kann durch Bedenken

vor falsch positiven CAD-Ergebnissen oder vor einem hohen Zeitaufwand in der Programmanwendung begründet sein [6].

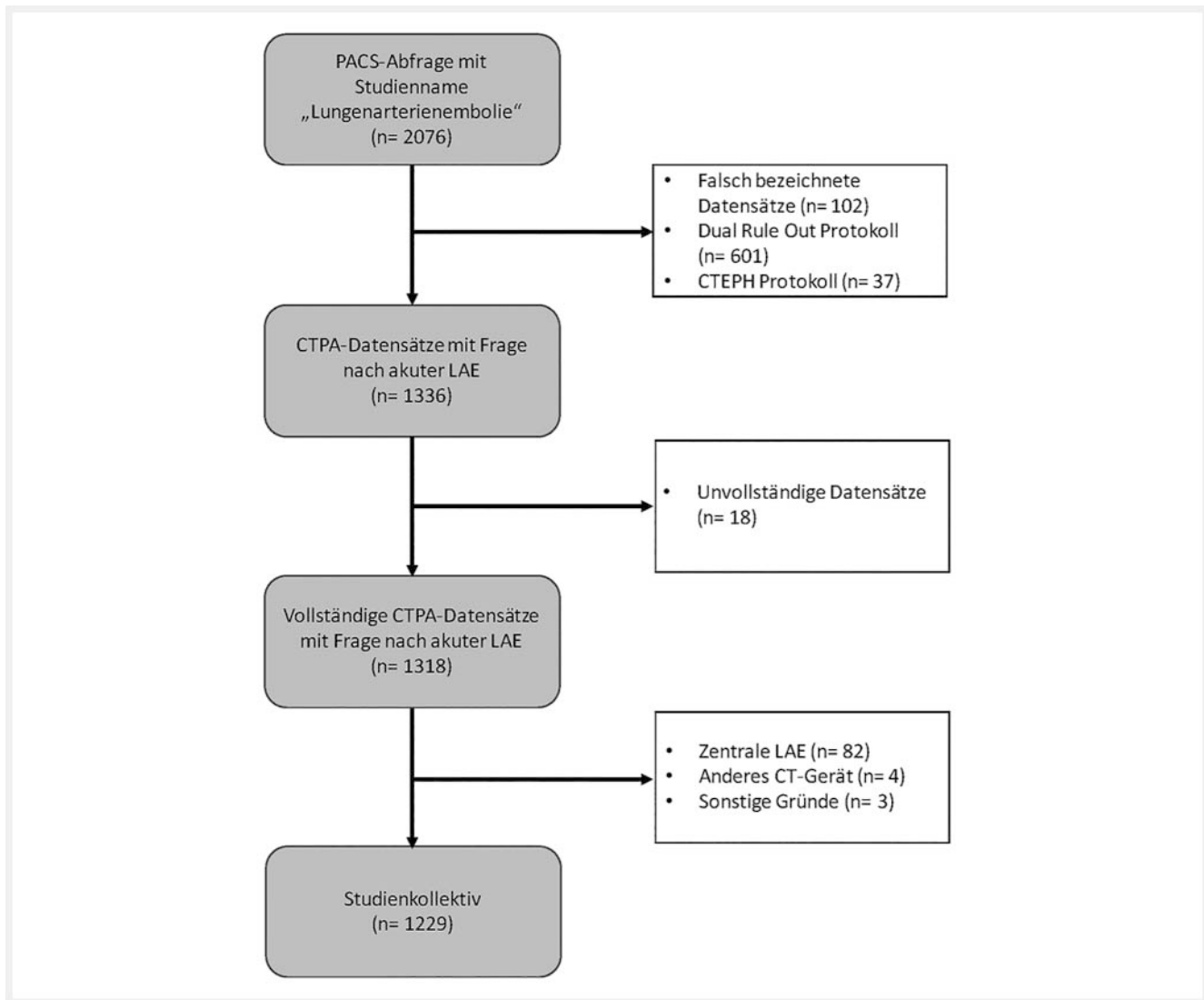
Das Ziel dieser Studie war die Evaluation eines kommerziellen, computerbasierten Algorithmus zur Detektion von Lungenarterienembolien anhand eines Patientenkollektivs einer großen deutschen Notaufnahme.

Material und Methoden

Diese monozentrische, retrospektive Studie wurde von der lokalen Ethikkommission genehmigt. Alle in die Studie eingeschlossenen CT-Untersuchungen waren klinisch indiziert.

Studienpopulation

In die Studie wurden alle Patient*innen eingeschlossen, die im Zeitraum vom 17.01.2014 bis zum 17.01.2017 eine CT-Untersuchung mit der Frage nach einer Lungenarterienembolie am Notfall-CT eines großen deutschen Universitätsklinikums erhalten hatten. Die entsprechenden Datensätze wurden über eine Stichwortsuche der Untersuchungsnamen im Picture Archiving and Communication System (PACS, Siemens Healthineers, Erlangen, Deutschland) ermittelt. Diese Recherche ergab für den genannten Zeitraum 2076 Datensätze. Das Flussdiagramm der ► **Abb. 1** zeigt, wie viele Datensätze aus welchen Gründen ausgeschlossen



► **Abb. 1** Flussdiagramm der in das Studienkollektiv eingeschlossenen Datensätze. PACS = Picture Archiving and Communication System; CTEPH = Chronisch-thromboembolische pulmonale Hypertonie; CTPA = Computertomografie-Pulmonalisangiografie; LAE = Lungenarterienembolie; CT = Computertomografie.

wurden. Ausgeschlossen wurden unvollständige sowie fälschlicherweise als CTPA deklarierte Datensätze; zudem Untersuchungen, die an einem anderen CT-Gerät, in Dual-Energy-Technik zum Ausschluss einer chronisch-thromboembolischen pulmonalen Hypertonie (CTEPH) oder als Kombinationsprotokoll („Dual-Rule-Out-Protokoll“) akquiriert wurden. Letztere konnten durch die CAD-Software nicht erkannt und somit nicht analysiert werden. Es wurde jeweils ein Datensatz bei Nachweis einer Knochenzementembolie, bei Zustand nach Pneumektomie und bei Zustand nach Fontan-Operation ausgeschlossen („sonstige Gründe“). Da zentrale Lungenarterienembolien in den Arteriae pulmonales eine Blickdiagnose darstellen, wurden diese Datensätze von der Studie ausgeschlossen. Das analysierte Studienkollektiv umfasste 1229 Patient*innen.

CT-Akquisition

Alle in das Studienkollektiv eingeschlossenen CTPA-Datensätze wurden am Dual-Source-SOMATOM-Force-CT (Siemens Healthineers, Erlangen, Deutschland) mit einer Gantry-Rotationszeit von 0,25 s/Rotation, einem Pitch von 1,9, einer Kollimation von $192 \times 0,6$ mm und einer Schichtdicke von 0,75 mm akquiriert. Röhrenspannung und Röhrenstrom wurden unter der Verwendung der Einstellungen „CAREkV“ und „CAREdose“ automatisch angepasst. Die Referenzpunkte waren hierbei eine Röhrenspannung von 100 kV und ein Röhrenstrom-Zeit-Produkt von 220 mAs. Zur Optimierung des Kontrasts wurde eine Bolus-Triggerung verwendet mit einer Region of Interest (ROI) im Truncus pulmonalis und einem automatisierten Scanbeginn 6 Sekunden nachdem die Dichte innerhalb der ROI 100 Hounsfield Units (HU) überschritten hatte. Die Kontrastmittelgabe erfolgte über eine Druckspritze mit einer Flussrate von 5 ml/s. Die Standardkontrastmittel-

gabe lag bei 50 ml, gefolgt von einem 100ml-Bolus 0,9%iger Kochsalzlösung. Von der genannten Kontrastmittelmenge wurde in 117 Fällen, bei einem dokumentierten BMI $\geq 30 \text{ kg/m}^2$, abgewichen. In diesen Fällen wurden 70 ml Kontrastmittel appliziert [7]. In den Jahren 2014 und 2015 wurde Imeron400 (Bracco Imaging GmbH, Konstanz, Deutschland) mit einer Jodkonzentration von 400 mg Jod/ml und in den Jahren 2016 und 2017 Ultravist370 (Bayer Vital GmbH, Leverkusen, Deutschland) mit einer Jodkonzentration von 370 mg Jod/ml verwendet.

Bildanalyse

CAD-Analyse

Die automatisierte Bildanalyse erfolgte mithilfe der Software *syngo.via* (Siemens Healthineers, Erlangen, Deutschland), Version VB20A im Rahmen der Studiauswertung. Für die LAE-CAD-Analyse wurden die Datensätze mit dem „CT Chest Pain + PE CAD“ geöffnet. In dem Modul „CT Lung Vessels“ erfolgte die automatisierte Analyse der pulmonalarteriellen Strombahn. LAE-CAD-Systeme nutzen hierfür eine Grauwertanalyse, um minderkontrastierte Areale in Angrenzung an kontrastierte Gefäßabschnitte zu detektieren [8]. Die automatisiert detektierten Ergebnisse, im Folgenden Findings genannt, müssen anschließend durch die anwendenden Radiolog*innen als richtig positiv und falsch positiv bewertet werden. Falsch positive Findings können gelöscht werden, richtig positive Findings werden bestätigt und können im Anschluss an die Auswertung ins PACS gesendet werden. Die Software selbst erstellt keine Wertung der einzelnen Findings oder des Gesamtbefundes; dies ist Aufgabe der anwendenden Radiolog*innen.

Die im Rahmen dieser Studie durchgeführte CAD-Analyse wurde durch die Goldstandardreader evaluiert. Alle CAD-Findings wurden im Abgleich mit der visuellen Analyse des Goldstandardreadings als richtig positiv oder falsch positiv gewertet. Für jedes falsch positive Ergebnis wurde in einem Consensus Reading die zugrunde liegende Ursache definiert: Weichteil, Pulmonalvene, Pneumonie, Atelektase, Durchmischungsphänomen, Drainagematerial, intrapulmonale Verkalkung.

Goldstandard

Die visuelle Bildanalyse im Rahmen des Goldstandardreadings erfolgte mithilfe der Software *syngo.via* (Siemens Healthineers, Erlangen, Deutschland), Version VB20A. Alle Datensätze des Studienkollektivs (n = 1229) wurden im Rahmen eines Consensus Readings von 2 Radiologen mit 5 und 15 Jahren Erfahrung in der Notfallradiologie visuell evaluiert. Dieses visuelle Consensus Reading wurde als Goldstandard definiert. Die Ergebnisse der CAD-Analyse wurden durch dieselben Radiologen mittels Consensus Reading evaluiert und im Vergleich zum Goldstandard als richtig positiv oder falsch positiv evaluiert. Ein Embolus, der durch das CAD mehrfach richtig markiert worden war, wurde als ein richtig positives Ergebnis gewertet. Jeder Embolus, der im Vergleich zum Goldstandard nicht durch das CAD markiert worden war, wurde als falsch negativ gewertet. War weder im Goldstandard noch im CAD eine LAE definiert worden, wurde die Untersuchung als richtig negativ gewertet.

$$\text{CNR} = \frac{(\text{HE}_{\text{Segmentarterie}} - \text{HE}_{\text{Muskulatur}})}{[(\text{STABW}_{\text{Segmentarterie}} - \text{STABW}_{\text{Muskulatur}})/2]}$$

► **Abb. 2** Formel zur Berechnung des Kontrast-Rausch-Verhältnisses. CNR = Contrast-to-noise ratio; HE = Hounsfield-Einheiten; STABW = Standardabweichung.

Für die Notfallbefundung einer CTPA ist es üblich, nicht jeden einzelnen Embolus, sondern die am weitesten proximal gelegenen Emboli und ihre Ausdehnung nach peripher zu beschreiben, da ein höherer Detailgrad keinen zusätzlichen Einfluss auf die Therapieentscheidung hat. Daher wurden die Notfallbefunde mit dem Goldstandard auf Patientenebene als richtig positiv, falsch positiv, richtig negativ und falsch negativ gewertet.

Routinebefundung Notaufnahme

Die in der Patientenakte dokumentierten radiologischen Befunde, die im Rahmen der Routinebefundung in der Notaufnahme erstellt wurden, wurden mit einer Siemens PACS-Software (Siemens Healthineers, Erlangen, Deutschland) visuell, ohne Unterstützung einer CAD-Software, analysiert.

Definition Lungenarterienembolie

Eine akute Lungenarterienembolie wurde als intravaskulärer, zentraler Füllungsdefekt mit umgebendem Kontrastmittelsaum oder als vollständige Verlegung eines fokal erweiterten Gefäßlumens definiert [9]. In Abhängigkeit der Lokalisation wurde die LAE als lobär, segmental oder subsegmental definiert und einer Seite zugeordnet. Ein Embolus, der kontinuierlich über mehrere anatomische Ebenen reichte, wurde der am weitesten zentral gelegenen Ebene zugeordnet und als ein Embolus gezählt [6].

Kontrast-Rausch-Verhältnis

Zur quantitativen Beurteilung der Bildqualität wurde das Kontrast-Rausch-Verhältnis (Contrast-to-noise ratio, CNR) in jedem Datensatz bestimmt, die Messungen wurden in *syngo.via* ermittelt. Die jeweils gleich großen ROIs wurden in das Lumen der superioren Unterlappensegmentarterie links sowie in die paravertebrale Muskulatur gelegt und die HU-Werte mit Standardabweichung (STABW) in die in ► **Abb. 2** dargestellte Formel eingesetzt. Im Falle eines Füllungsdefektes in der apikalen Unterlappensegmentarterie links wurde die ROI in eine angrenzende segmentale Unterlappenarterie ohne Füllungsdefekt gelegt.

Statistische Analyse

Stetige Daten wurden als Mittelwert mit Standardabweichung sowie in Prozent angegeben. Kategoriale Daten wurden mit absoluter sowie relativer Häufigkeit (in Prozent) angegeben. Sensitivität und Spezifität werden in Prozent mit dem korrespondierenden 95%-Konfidenzintervall (KI) aufgeführt. Die Abhängigkeit zwischen dem Auftreten segmentaler und subsegmentaler Lungenarterienemboli wurde mit dem Chi-Quadrat-Test geprüft. Die

Korrelation zwischen der Anzahl der CAD-Ergebnisse und der CNR wurde mit der Korrelationsanalyse nach Spearman untersucht. Ein p -Wert $< 0,05$ wurde als Signifikanzniveau festgelegt. Für die Auswertung wurden Excel 2016, Version 16.15 (Microsoft Corporation, Redmond, WA, USA) und Prism 8, Version 8.0.0 (GraphPad Software Inc., La Jolla, CA, USA) verwendet.

Ergebnisse

Patientencharakteristiken

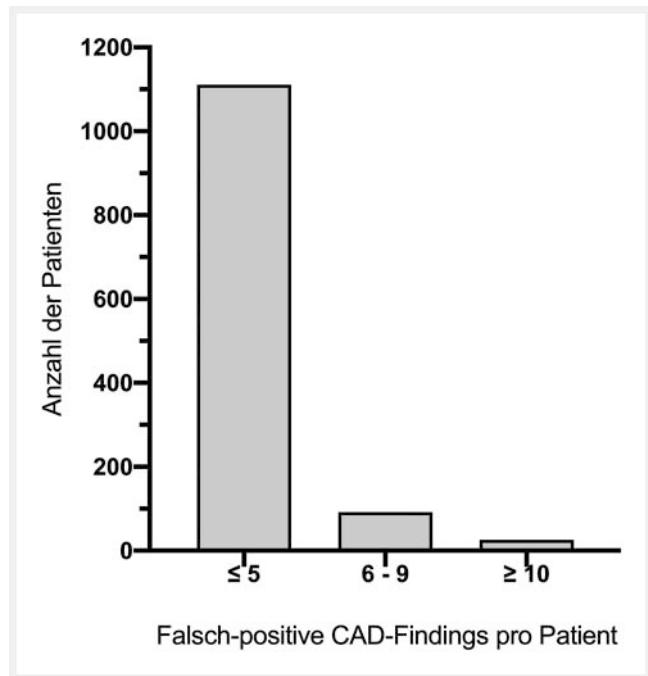
Die Datensätze von 1229 Patient*innen wurden in die Studie eingeschlossen, hierunter waren 607 (49,4%) männliche und 622 (50,6%) weibliche Patient*innen im Alter von 10–97 Jahren (Mittelwert 64 ± 17 Jahre). Hierbei divergierte das Studienkollektiv nicht wesentlich vom Kollektiv der initialen Datenbankabfrage ($n = 2076$) mit 1062 (51,2%) männlichen und 1014 (48,8%) weiblichen Patient*innen im Alter zwischen 0 und 97 Jahren (Mittelwert 68 ± 18 Jahre).

Ergebnisse Goldstandard

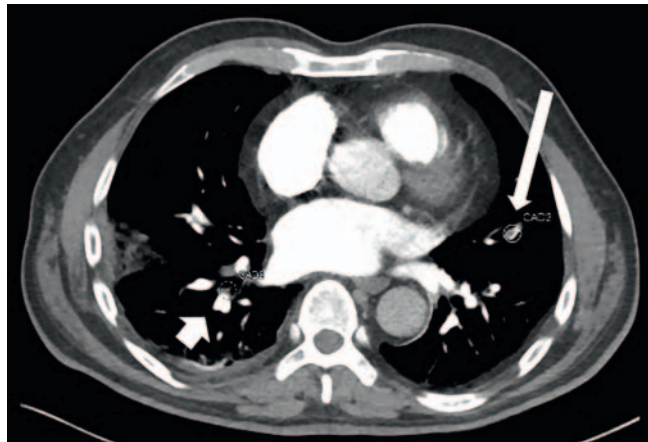
Bei 182 von 1229 (14,8%) Patient*innen wurde mindestens eine Lungenarterienembolie detektiert, hiervon waren 93 (7,57%) männliche und 89 (7,24%) weibliche Patient*innen. Bei 106 (58,24%) Patient*innen wurden multiple Embolien, bei 76 (41,76%) Patient*innen wurde eine singuläre Lungenarterienembolie detektiert. In der Summe wurden 504 Embolien beschrieben, hiervon 51 (10,12%) auf Lappen-, 219 (43,45%) auf Segment- und 234 (46,43%) auf Subsegmentebene. Der überwiegende Anteil der Embolien wurde rechts segmental (137 Embolien; 27,18%) und rechts subsegmental (146 Embolien; 28,97%) nachgewiesen. Der Chi-Quadrat-Test bestätigte eine statistische Abhängigkeit zwischen dem Auftreten segmentaler und subsegmentaler Lungenarterienembolien ($\chi^2(1, n = 453) = 28,2; p < 0,0001$).

Ergebnisse CAD

Die CAD-Analyse ergab bei den 1229 Datensätzen insgesamt 3331 Findings. Diese reichten von 0 bis maximal 25 Findings pro Datensatz mit einem Mittelwert von $2,7 \pm 2,61$ Findings. Von 182 Patient*innen mit Nachweis einer LAE gemäß Goldstandard hat die CAD-Analyse bei 118 Patient*innen mindestens ein richtig positives Finding angezeigt, gemäß einer Sensitivität von 64,8% (95%-KI 0,58–0,72). Bei 64 Patient*innen wurden die durch das Goldstandardreading markierten Embolien durch die CAD-Software nicht markiert. In 13 dieser 64 Datensätze wurde überhaupt kein Finding detektiert. In den übrigen 51 Datensätzen wurden Findings markiert, hierbei handelte es sich jedoch um falsch positive Findings. Von 1047 Patient*innen, die gemäß Goldstandard keine LAE aufwiesen, hat die CAD-Analyse bei 821 Patient*innen mindestens ein Finding und bei 226 Patient*innen kein Finding angezeigt, einer Spezifität von 21,6% (95%-KI 0,19–0,24) sowie einem positiven prädiktiven Wert (PPW) von 12,6% (95%-KI 0,11–0,15) und einem negativen prädiktiven Wert (NPW) von 77,9% (95%-KI 0,73–0,82) entsprechend. Die Häufigkeitsverteilung der Anzahl der falsch positiven CAD-Findings pro Patient*in ist in ► **Abb. 3** dargestellt.



► **Abb. 3** Säulendiagramm der Häufigkeitsverteilung der falsch positiven CAD-Findings. CAD = computerassistierter Detektionsalgorithmus.



► **Abb. 4** Beispiele für ein richtig positives (langer, weißer Pfeil) und ein falsch positives (kurzer, weißer Pfeil) Finding.

Von insgesamt 504 im Goldstandard definierten Embolien hat die CAD 258 Embolien detektiert (► **Abb. 4**), einer Sensitivität von 51,2% (95%-KI 0,47–0,56) entsprechend. Auf Lappenebene wurden hierbei 24 von 51 Embolien (Sensitivität: 47,1%; 95%-KI 0,33–0,61), auf Segmentebene 118 von 219 Embolien (Sensitivität: 53,9%; 95%-KI 0,47–0,61) und auf Subsegmentebene 116 von 234 Embolien (Sensitivität: 49,6%; 95%-KI: 0,43–0,56) detektiert.

Von den 3331 CAD-Findings waren 3073 (92,3%) als falsch positiv durch das Consensus Reading gewertet worden (► **Abb. 4**). Im Mittel wurden bei jedem Patienten somit $2,5 \pm 2,54$ falsch positive Findings angezeigt. Jedem der falsch positiven CAD-Befunde wurde

im Rahmen des Consensus Readings eine zugrunde liegende Ursache zugeordnet, die in ► **Tab. 1** zusammengefasst sind.

Der Chi-Quadrat-Test betätigte einen statistischen Zusammenhang für die Anzahl der CAD-Findings und dem LAE-Status, ($\chi^2(1, n = 1229) = 20,6; p < 0,00001$), die zugrunde liegende Kontingenztafel mit Angabe der Spaltenprozentage ist in ► **Tab. 2** dargestellt.

Ergebnisse Routinebefundung Notaufnahme

Im Rahmen der Routinebefundung durch die diensthabenden Radiolog*innen wurde bei 180 der insgesamt 182 durch den Goldstandard definierten LAE-positiven Datensätze eine Lungenarterienembolie beschrieben. Dies entspricht einer Sensitivität der Routinebefundung in der Notaufnahme im Vergleich zum Goldstandardreading von 98,9% (95%-KI 0,97–1,00). In beiden Fällen wurde durch das Goldstandardreading eine singuläre subsegmentale Lungenarterienembolie detektiert. Die im Rahmen der Studie durchgeführte CAD-Analyse zeigte nur in einem der beiden Fälle den durch das Goldstandardreading definierten Embolus als Finding an. Im zweiten Fall wurde der durch das Goldstandardreading definierte Embolus durch die CAD-Software nicht markiert. Im Vergleich zum Goldstandardreading wurde kein im Rahmen der Routinebefundung der Notaufnahme erstellter Befund als falsch positiv gewertet.

Dauer CAD-Analyse

In einer Stichprobe von 50 Datensätzen wurde die Dauer analysiert, die der LAE-CAD der syngo.via-Software vom Öffnen des Datensatzes bis zur Anzeige der CAD-Findings benötigte; diese lag bei $92,3 \pm 15,5$ Sekunden. Die für die Evaluation der CAD-Findings benötigte Zeit variierte in Abhängigkeit der Anzahl der CAD-Findings von 0 Sekunden, wenn keine Findings angezeigt wurden, bis zu 92 Sekunden im Falle des Maximums von 25 CAD-Findings. Im Mittel dauerte die Bewertung der CAD-Findings $11,76 \pm 16,27$ Sekunden.

Kontrast-Rausch-Verhältnis

Die CNR lag zwischen 2,8 und 91,8 mit einem Mittelwert von $24,5 \pm 12,6$. Die errechneten CNR-Werte im Verhältnis zur Anzahl der Datensätze des Studienkollektivs sind in ► **Abb. 5** dargestellt. Die Spearman-Korrelation zwischen der CNR und der Anzahl der falsch positiven CAD-Findings ($r_s = 0,085$) zeigte keine Korrelation der Variablen.

Diskussion

Mit zunehmender Verfügbarkeit und technischer Verbesserung der CT resultiert ein vermehrter Einsatz der CTPA in der Lungenarterienemboliediagnostik [4]. Eine Herausforderung stellen hierbei die peripheren segmentalen sowie die subsegmentalen LAE dar, die insbesondere bei isoliertem Auftreten übersehen werden können [6]. Die Relevanz einer singulären, subsegmentalen LAE (ssLAE) wird in der Literatur kontrovers diskutiert. Während manche Studien eine Überdiagnostik mit konsekutiver Übertherapie der ssLAE

► **Tab. 1** Übersicht zugrundeliegender Ursachen für falsch positive CAD-Findings als absolute Werte und in Prozent. CAD – computer-assistierter Detektionsalgorithmus..

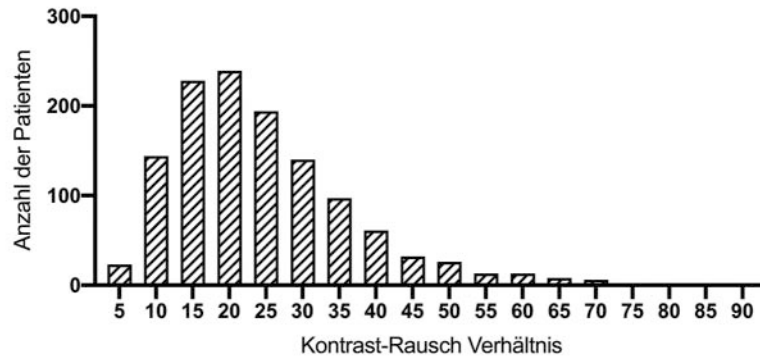
Ursache Falsch positiver CAD-Findings	Anzahl Falsch positiver CAD-Findings
Weichteil	1876 (61 %)
Pulmonalvene	742 (24,1 %)
Pneumonie	357 (11,6 %)
Atelektase	59 (1,9 %)
Durchmischungsphänomen	29 (0,9 %)
Drainagematerial	9 (0,3 %)
intrapulmonale Verkalkung	1 (0,03 %)

► **Tab. 2** Vierfeldertafel der Goldstandardbefunde und der CAD-Findings auf Patientenebene. Goldst. – Goldstandard; LAE – Lungenarterienembolie; CAD – computer-assistierter Detektionsalgorithmus.

	Goldst. Nachweis LAE	Goldst. kein Nachweis LAE	gesamt
mind. 1 CAD-Finding	169 (92,9 %)	821 (78,4 %)	990
kein CAD-Finding	13 (7,1 %)	226 (21,6 %)	239
gesamt	182	1047	1229

beschrieben haben, schlussfolgern Autoren anderer Studien, dass der klinische Outcome von Patient*innen mit subsegmentaler LAE und jenen mit weiter proximal nachgewiesenen LAE vergleichbar sei [10–12]. Aufgrund der uneindeutigen Datenlage empfiehlt die geltende Leitlinie der Europäischen Gesellschaft für Kardiologie (European Society for Cardiology, ESC) eine individuelle Therapieentscheidung bei subsegmentalen LAE in Abhängigkeit der patienteneigenen Risikofaktoren für eine LAE und Blutungskomplikation [13]. Somit bleibt die Detektion subsegmentaler LAE in der CTPA weiterhin hoch relevant. Studien zufolge können CAD-Systeme bei der Analyse von CTPA-Datensätzen unterstützen. In bisherigen Studien wurden jedoch insbesondere kleine Patientenkollektive oder CAD-Prototypen evaluiert [6, 14–16].

Nach unserem Kenntnisstand ist dies die erste Studie, die ein kommerzielles CAD-System anhand eines großen Patientenkollektivs eines Klinikums der Maximalversorgung in Deutschland untersucht und hierbei die CAD-Analyse sowohl mit den Befunden der Notfallversorgung als auch mit dem Expertenreading verglichen hat. Im Rahmen unserer Studie wurden 1229 CTPA-Datensätze von 2 Radiologen als Goldstandardreading visuell analysiert. Es wurden 504 Emboli in 182 Datensätzen detektiert und als Goldstandard für die Bewertung des CAD-Systems definiert. Das LAE-CAD detektierte hiervon 258 Emboli, einer Sensitivität von



► **Abb. 5** Säulendiagramm der Häufigkeitsverteilung der Werte des Kontrast-Rausch-Verhältnisses.

51,2% auf Embolusebene entsprechend. Hierbei lag die Sensitivität der Detektion segmentaler Emboli mit 53,9% am höchsten. Die CAD-Software zeigte bei 118 Patienten von insgesamt 182 Patienten mit LAE mindestens ein richtig positives Finding an, einer Sensitivität von 64,8% auf Patientenebene entsprechend. Diese Werte liegen hierbei unterhalb der in der Literatur beschriebenen Werte, die in den letzten Studien zwischen 55,4 und 76% Sensitivität auf Embolusebene und zwischen 83 und 100% auf Patientenebene lagen [14–19]. Im Rahmen unserer Studienauswertung wurden zudem alle Routinebefunde der Notaufnahme mit dem Goldstandardreading verglichen. Abweichungen konnten in nur 2 Fällen nachgewiesen werden; in diesen wurde durch das Goldstandardreading jeweils ein subsegmentaler Embolus definiert, der im klinischen Befund nicht dokumentiert worden war. Die CAD-Software hatte nur in einem der beiden Fälle den Embolus markiert, im zweiten Fall hätte auch eine ergänzend durchgeführte CAD-Analyse den falsch negativen Befund in der klinischen Routine nicht verhindern können.

Bei Patienten mit LAE-Nachweis wurde mit höherer Wahrscheinlichkeit mindestens ein Finding angezeigt. Im Mittel hat die CAD-Analyse $2,7 \pm 2,61$ Findings insgesamt und $2,5 \pm 2,54$ falsch positive Findings pro Patient*in angezeigt. Diese Werte liegen deutlich unter den Werten anderer LAE-CAD-Studien, die bis zu einem maximalen Mittelwert von 14,4 falsch positiven Findings pro Patient*in reichen [6, 16–18] und gering oberhalb der Studien von Tajbakhsh et al. und Lahiji et al. mit einem Mittelwert von nur 2 bzw. 2,1 falsch positiven Findings pro Datensatz [14, 15]. Das Patientenkollektiv allein kann hierbei nicht für die im Rahmen unserer Studie deutlich niedrigere Sensitivität herangezogen werden, da auch Wittenberg et al. ein Patientenkollektiv einer universitären Notaufnahme analysiert haben und eine Sensitivität von 95% auf Patientenebene berichtet haben [16]. Zudem wäre bei einem Kollektiv, das aus Intensiv- und Notaufnahme patient*innen zusammengesetzt ist, primär eine erhöhte Anzahl falsch positiver Befunde zu erwarten, da die Symptome in einer Vielzahl der Fälle durch andere Ursachen, wie beispielsweise Pneumonien oder Atelektasen, erklärt werden können. Es ist somit zu vermuten, dass der im Vergleich zu anderen Studien niedrigen durchschnittlichen Rate falsch positiver Findings in unserem Fall mit einer hohen Rate an falsch negativen Befunden Rechnung getragen wird. In Übereinstimmung mit unseren Ergeb-

nissen werden Anschnitte von mediastinalen und hilären Weichteilen sowie von Pulmonalvenen als die häufigsten Ursachen falsch positiver CAD-Findings angegeben [6, 15, 16].

Einen signifikanten Zusammenhang zwischen der Bildqualität und der Anzahl falsch positiver Befunde konnten wir, im Gegensatz zu anderen Studien, nicht nachweisen. Hierbei muss jedoch bedacht werden, dass in der vorliegenden Studie die Bildqualität anhand der CNR ausschließlich quantitativ und nicht semiquantitativ anhand einer Likert-Skala bewertet wurde [6, 17, 20].

Bei der Verwendung zusätzlicher Analysesoftware ist der zeitliche Aufwand, der hierdurch entsteht, ein relevantes Qualitätskriterium. Dies gilt insbesondere in der Notfallbefundung. Wir haben daher die zusätzlich für das LAE-CAD benötigte Zeit in einer Stichprobe analysiert. Die Software zeigte die Findings im Mittel nach $92,3 \pm 15,5$ Sekunden an, im Mittel wurden dann weitere $11,76 \pm 16,27$ Sekunden für die Evaluation der CAD-Findings benötigt. Das Einladen und die Analyse des Datensatzes durch die CAD-Software beansprucht somit deutlich mehr Zeit als die Evaluation der Findings durch die Radiolog*innen. Unser CAD benötigt hierbei im Vergleich zu Wittenberg et al. im Durchschnitt 3-mal so lange für die Analyse [16]. Da dieser Prozess in der klinischen Anwendung jedoch parallel zur visuellen Analyse abläuft, wird hierdurch keine zusätzliche Zeit beansprucht. Entgegen der allgemeinen Wahrnehmung ist die Evaluation der CAD-Findings in der überwiegenden Anzahl der Fälle innerhalb weniger Sekunden, im Durchschnitt nach $11,76 \pm 16,27$ Sekunden, beendet. In wenigen Fällen wurde eine Dauer von bis zu 92 Sekunden benötigt, um die maximale Anzahl an 25 CAD-Findings zu evaluieren.

Die Entwicklungen der letzten Jahre haben verdeutlicht, dass öffentlich zugängliche Datensätze benötigt werden, um LAE-CAD-Systeme untereinander vergleichen zu können [21]. Zudem haben Studien gezeigt, dass das Innovationspotenzial noch nicht ausgeschöpft ist. So wurden monoenergetische Rekonstruktionen von Dual-Energy-Datensätzen für die CAD-Analyse verwendet, was zu einer erhöhten Sensitivität führte [22, 23] sowie die 3-dimensionale Segmentierung der pulmonalarteriellen Strombahn optimiert [24, 25]. Diese Entwicklungen deuten an, dass mit weiterer Optimierung das LAE-CAD Bestandteil unseres klinischen Alltags werden kann.

Als Limitation unserer Studie ist zum einen das monozentrische, retrospektive Studiendesign zu nennen. Zum anderen mussten 601 Datensätze ausgeschlossen werden, die zum gleichzeitigen Ausschluss einer LAE und einer Aortendissektion als Dual-Rule-Out-Protokoll akquiriert wurden. Diese Datensätze konnten nicht in das LAE-CAD eingeladen werden, sodass ein Selektionsbias möglich ist.

Schlussfolgerung

Die Auswertung des kommerziellen LAE-CAD-Systems anhand eines großen Patientenkollektivs der Notfallversorgung zeigte den Optimierungsbedarf der Software insbesondere in Bezug auf die falsch negativen Befunde auf. Es verdeutlicht die Rolle als unterstützende Analysesoftware, die die visuelle Evaluation durch die Radiolog*innen (noch) nicht ersetzen kann.

KLINISCHE RELEVANZ DER STUDIE

- Da die LAE eine häufige Erkrankung mit potenziell hoher Morbidität und Mortalität darstellt, ist die CTPA eine häufig durchgeführte Untersuchung im Notfallbetrieb.
- Insbesondere singuläre subsegmentale LAE können übersehen werden.
- LAE-CAD-Systeme können die Radiolog*innen bei der Analyse ohne wesentlich erhöhten Zeitaufwand unterstützen.
- Bei der Anwendung muss bedacht werden, dass die CAD-Analyse sowohl falsch negative als auch falsch positive Befunde generiert und die visuelle Evaluation nicht ersetzen kann.

Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

- [1] Calder KK, Herbert M, Henderson SO. The mortality of untreated pulmonary embolism in emergency department patients. *Ann Emerg Med* 2005; 45: 302–310. doi:10.1016/j.annemergmed.2004.10.001
- [2] Wendelboe AM, Raskob GE. Global Burden of Thrombosis: Epidemiologic Aspects. *Circ Res* 2016; 118: 1340–1347. doi:10.1161/circresaha.115.306841
- [3] Diehm C, Noppeney T, Nüllen H. Epidemiologie der venösen Thromboembolie. *Gefäßchirurgie* 2012; 17: 275–279. doi:10.1007/s00772-011-0929-8
- [4] Sharma S, Lucas CD. Increasing use of CTPA for the investigation of suspected pulmonary embolism. *Postgrad Med* 2017; 129: 193–197. doi:10.1080/00325481.2017.1281084
- [5] Schissler A, Rozenshtein A, Kulon M et al. CT Pulmonary Angiography: Increasingly Diagnosing Less Severe Pulmonary Emboli. *PloS one* 2013; 8: e65669 doi:10.1371/journal.pone.0065669
- [6] Kligerman SJ, Lahiji K, Galvin JR et al. Missed pulmonary emboli on CT angiography: assessment with pulmonary embolism-computer-aided detection. *Am J Roentgenol* 2014; 202: 65–73. doi:10.2214/ajr.13.11049
- [7] Hawley PC, Hawley MP. Difficulties in diagnosing pulmonary embolism in the obese patient: a literature review. *Vasc Med* 2011; 16: 444–451. doi:10.1177/1358863x11422571
- [8] Buhmann S, Herzog P, Liang J et al. Clinical evaluation of a computer-aided diagnosis (CAD) prototype for the detection of pulmonary embolism. *Acad Radiol* 2007; 14: 651–658. doi:10.1016/j.acra.2007.02.007
- [9] Moore AJE, Wachsmann J, Chamarthy MR et al. Imaging of acute pulmonary embolism: an update. *Cardiovasc Diagn Ther* 2018; 8: 225–243. doi:10.21037/cdt.2017.12.01
- [10] Eyer BA, Goodman LR, Washington L. Clinicians' response to radiologists' reports of isolated subsegmental pulmonary embolism or inconclusive interpretation of pulmonary embolism using MDCT. *Am J Roentgenol* 2005; 184: 623–628. doi:10.2214/ajr.184.2.01840623
- [11] Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. *Arch Intern Med* 2011; 171: 831–837. doi:10.1001/archinternmed.2011.178
- [12] den Exter PL, van Es J, Klok FA et al. Risk profile and clinical outcome of symptomatic subsegmental acute pulmonary embolism. *Blood* 2013; 122: 1144–1149; quiz 1329. doi:10.1182/blood-2013-04-497545
- [13] Konstantinides SV, Meyer G, Becattini C et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). *European Heart Journal* 2019; 41: 543–603. doi:10.1093/eurheartj/ehz405
- [14] Tajbakhsh N, Gotway MB, Liang J. Computer-Aided Pulmonary Embolism Detection Using a Novel Vessel-Aligned Multi-planar Image Representation and Convolutional Neural Networks. Cham: Springer International Publishing; 2015: 62–69
- [15] Lahiji K, Kligerman S, Jeudy J et al. Improved accuracy of pulmonary embolism computer-aided detection using iterative reconstruction compared with filtered back projection. *Am J Roentgenol* 2014; 203: 763–771. doi:10.2214/ajr.13.11838
- [16] Wittenberg R, Peters JF, Sonnemans JJ et al. Computer-assisted detection of pulmonary embolism: evaluation of pulmonary CT angiograms performed in an on-call setting. *Eur Radiol* 2010; 20: 801–806. doi:10.1007/s00330-009-1628-7
- [17] Wittenberg R, Peters JF, Weber M et al. Stand-alone performance of a computer-assisted detection prototype for detection of acute pulmonary embolism: a multi-institutional comparison. *Br J Radiol* 2012; 85: 758–764. doi:10.1259/bjr/26769569
- [18] Özkan H, Osman O, Şahin S et al. A novel method for pulmonary embolism detection in CTA images. *Comput Methods Programs Biomed* 2014; 113: 757–766. doi:10.1016/j.cmpb.2013.12.014
- [19] Krissak R, Henzler T, Reichert M et al. Enhanced visualization of lung vessels for diagnosis of pulmonary embolism using dual energy CT angiography. *Invest Radiol* 2010; 45: 341–346. doi:10.1097/RLI.0b013e3181dfda37
- [20] Wittenberg R, Peters JF, Sonnemans JJ et al. Impact of Image Quality on the Performance of Computer-Aided Detection of Pulmonary Embolism. *American Journal of Roentgenology* 2011; 196: 95–101. doi:10.2214/Am J Roentgenol.09.4165
- [21] Masoudi M, Pourreza HR, Saadatmand-Tarzan M et al. A new dataset of computed-tomography angiography images for computer-aided detection of pulmonary embolism. *Sci Data* 2018; 5: 180180 doi:10.1038/sdata.2018.180

- [22] Ma G, Dou Y, Dang S et al. Influence of Monoenergetic Images at Different Energy Levels in Dual-Energy Spectral CT on the Accuracy of Computer-Aided Detection for Pulmonary Embolism. *Acad Radiol* 2019; 26: 967–973. doi:10.1016/j.acra.2018.09.007
- [23] Kröger JR, Hicethier T, Pahn G et al. Influence of spectral detector CT based monoenergetic images on the computer-aided detection of pulmonary artery embolism. *Eur J Radiol* 2017; 95: 242–248. doi:10.1016/j.ejrad.2017.08.034
- [24] Tajbakhsh N, Shin JY, Gotway MB et al. Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation. *Med Image Anal* 2019; 58: 101541 doi:10.1016/j.media.2019.101541
- [25] Zhang C, Sun M, Wei Y et al. Automatic segmentation of arterial tree from 3D computed tomographic pulmonary angiography (CTPA) scans. *Comput Assist Surg (Abingdon)* 2019; 24: 79–86. doi:10.1080/24699322.2019.1649077