

Accuracy of Chest CT for Differentiating COVID-19 from COVID-19 Mimics

Diagnostische Genauigkeit des Thorax-CT zur Unterscheidung von COVID-19-Pneumonie und COVID-19-Mimics

Authors

Marwin-Jonathan Sähn¹ , Can Yüksel¹ , Sebastian Keil¹, Marcel P. Zeisberger¹, Manuel Post¹ , Michael Kleines², Jörg Christian Brokmann³, Christian Hübel³, Christiane K. Kuhl¹, Peter Isfort¹, Maximilian Franz Schulze-Hagen¹ 

Affiliations

- 1 Department of Diagnostic and Interventional Radiology, Universitätsklinikum Aachen, Germany
- 2 Laboratory Diagnostics Center, Universitätsklinikum Aachen, Germany
- 3 Emergency Department, Universitätsklinikum Aachen, Germany

Key words

COVID-19, diagnostic accuracy, CT, atypical pneumonia

received 29.09.2020

accepted 19.01.2021

published online 26.03.2021

Bibliography

Fortschr Röntgenstr 2021; 193: 1081–1091

DOI 10.1055/a-1388-7950

ISSN 1438-9029

© 2021, Thieme. All rights reserved.

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

Correspondence

Marwin-Jonathan Sähn

Interventional and diagnostic Radiology, Uniklinik RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany

Tel.: ++49/241/8 08 85 19

msaehn@ukaachen.de

ZUSAMMENFASSUNG

Ziel Bestimmung der Leseleistung von Radiologen mit unterschiedlichen Fachkenntnissen hinsichtlich der Unterscheidung von COVID-19 von anderen atypischen Pneumonien. Als Limitierung der Thorax-CT in der Identifizierung von Patienten mit COVID-19 wird eine geringe Spezifität in der Unterscheidung von COVID-19 von anderen atypischen Pneumonien („COVID-19-Mimics“) beschrieben. Inzwischen hat sich das Verständnis der morphologischen Muster von COVID-19 verbessert und scheint relativ spezifisch zu sein.

Material und Methoden Im Zeitraum von Februar bis April 2020 wurden 60 Patienten mit COVID-19-Pneumonie mittels Thorax-CT in unserem Hause untersucht. Die Fälle wurden

einer vergleichbaren Kontrollgruppe mit ähnlicher Geschlechterverteilung, Alter und Vorerkrankungen gegenübergestellt, die eine CT-Thorax bei atypischer Pneumonie vor Januar 2020 erhielt. Eingeschlossen wurden andere virale, Pilz- und atypische bakterielle Erreger. Alle 120 Fälle wurden verblindet von 2 radiologischen Fachärzten und 2 Assistenzärzten hinsichtlich der Wahrscheinlichkeit einer COVID-19-Pneumonie anhand des COV-RADS-Score beurteilt. Die Ergebnisse wurden mittels Clopper-Pearson-95 %-Konfidenzintervallen, Youden-Index für die Testgütekriterien und Fleiss' Kappa ausgewertet.

Ergebnisse Insgesamt erkannten die Radiologen das Vorliegen einer COVID-19-Pneumonie in 219/240 Wertungen (Sensitivität: 91 %; 95 %-KI 86,9–94,5 %) und das eines „COVID-19-Mimics“ in 159/240 Wertungen (Spezifität: 66,3 %; 95 %-KI 59,9 %–72,2 %). Dies entspricht einer diagnostischen Genauigkeit von 78,8 % (378/480 Wertungen; 74,8–82,3 %). Individuelle diagnostische Genauigkeiten reichten von 74,2 % (89/120) bis 84,2 % (101/120) und korrelierten nicht signifikant mit der Berufserfahrung. Der Youden-Index betrug 0,57. Die Übereinstimmung der Radiologen war moderat ($\kappa = 0,53$).

Zusammenfassung In dieser mit atypischen Pneumonien angereicherten Kohorte konnten die Radiologen anhand der CT-Untersuchung COVID-19-Pneumonien von „COVID-Mimics“ mit moderater diagnostischer Genauigkeit unterscheiden. Hierbei zeigte die Berufserfahrung der Radiologen keinen direkten Einfluss auf die Ergebnisse.

Kernaussagen:

- Eine Unterscheidung zwischen COVID-19- und anderen atypischen Pneumonien („COVID-Mimics“) in der CT ist im Szenario des direkten Vergleichs (keine Negativbefunde) mit moderater diagnostischer Genauigkeit möglich.
- Die Berufserfahrung hatte keinen direkten Einfluss auf die Ergebnisse.
- Trotz der ähnlichen Verteilung von Infiltraten konnten die Radiologen anhand der COV-RADS-Klassifikation die Wahrscheinlichkeiten für das Vorliegen einer COVID-Pneumonie reliabel und standardisiert im größeren Anteil der Fälle einschätzen.

ABSTRACT

Purpose To determine the performance of radiologists with different levels of expertise regarding the differentiation of COVID-19 from other atypical pneumonias. Chest CT to identify patients suffering from COVID-19 has been reported to be limited by its low specificity for distinguishing COVID-19 from other atypical pneumonias (“COVID-19 mimics”). Meanwhile, the understanding of the morphologic patterns of COVID-19 has improved and they appear to be fairly specific.

Materials and Methods Between 02/2020 and 04/2020, 60 patients with COVID-19 pneumonia underwent chest CT in our department. Cases were matched with a comparable control group of 60 patients of similar age, sex, and comorbidities, who underwent chest CT prior to 01/2020 for atypical pneumonia caused by other pathogens. Included were other viral, fungal, and bacterial pathogens. All 120 cases were blinded to patient history and were reviewed independently by two radiologists and two radiology residents. Readers rated the probability of COVID-19 pneumonia according to the COV-RADS classification system. Results were analyzed using Clopper-Pearson 95% confidence intervals, Youden's Index for test quality criteria, and Fleiss' kappa statistics.

Results Overall, readers were able to correctly identify the presence of COVID-19 pneumonia in 219/240 (sensitivity: 91%; 95%-CI: 86.9%–94.5%), and to correctly attribute CT findings to COVID-19 mimics in 159/240 ratings (specificity: 66.3%; 59.9%–72.2%), yielding an overall diagnostic accuracy

of 78.8% (378/480; 74.8%–82.3%). Individual reader accuracy ranged from 74.2% (89/120) to 84.2% (101/120) and did not correlate significantly with reader expertise. Youden's Index was 0.57. Between-reader agreement was moderate ($\kappa = 0.53$).

Conclusion In this enriched cohort, radiologists were able to distinguish COVID-19 from “COVID-19 mimics” with moderate diagnostic accuracy. Accuracy did not correlate with reader expertise.

Key Points:

- In a scenario of direct comparison (no negative findings), CT allows the differentiation of COVID-19 from other atypical pneumonias (“COVID mimics”) with moderate accuracy.
- Reader expertise did not significantly influence these results.
- Despite similar patterns and distributions of pulmonary findings, radiologists were able to estimate the probability of COVID-19 pneumonia using the COV-RADS classification in a standardized manner in the larger proportion of cases.

Citation Format

- Sähn M, Yüksel C, Keil S et al. Accuracy of Chest CT for Differentiating COVID-19 from COVID-19 Mimics. *Fortschr Röntgenstr* 2021; 193: 1081–1091

Introduction

Chest computed tomography (CT) is a useful diagnostic tool to help identify COVID-19-associated (coronavirus disease 2019) pneumonia [1–7]. Typical imaging patterns of COVID-19, such as focal ground-glass opacities, interlobular thickening, crazy-paving pattern, consolidations, and a tendency toward peripheral and basal distribution of lesions, have recently been described [4–11]. Sensitivity rates for COVID-19 on chest CT scans have been shown to be high, ranging from 90% to 98% in recent publications [3, 12–14].

However, the specificity with which chest CT can help establish the diagnosis of COVID-19 has been reported to be insufficient. This is why the American College of Radiology (ACR) discourages the use of CT as a first-line test in its recent publication and suggests conservative use in “symptomatic patients with specific clinical indications” [15]. The ACR argues that other atypical pneumonias, such as influenza A and B, SARS-1 and MERS, and other non-infectious pathologies, such as drug-induced pneumonitis, show overlapping morphologic patterns and might therefore mimic COVID-19-associated pneumonia in radiological imaging [16–18]. Indeed, one of the first studies on the systematic use of chest CT for the diagnosis of COVID-19 reported a specificity as low as 25% [3]. Since this study, the features of COVID-19-associated pneumonia have been further established. More recent studies found much higher specificities of chest CT to exclude

COVID-19 of 91–100% [14, 20, 21] – at sensitivity levels that ranged from 72% to 94%.

Accordingly, to further investigate the accuracy with which chest CT can help distinguish COVID-19-associated pneumonia from other atypical pneumonias, and to determine whether reader expertise would drive reader performance, we conducted a study on an enriched cohort selected to exclusively include patients suffering from atypical pneumonia either due to COVID-19 or other causes.

Materials and Methods

This retrospective study was approved by the local ethics committee (EK 097/20). Using a query in our university hospital's database, we included the first 60 patients with COVID-19 and chest CTs conducted within seven days of the initial positive PCR result. These patients were treated in the time period of late 02/2020 to early 04/2020. All COVID-19 cases were confirmed by in-house reverse transcriptase polymerase chain reaction (RT-PCR) testing from nasopharyngeal or oral swabs using test kits by Altona diagnostics (Hamburg, Germany).

Atypical pneumonias were selected using a standardized query for all patients with ICD-10 J09–J18 (influenza and pneumonia) as well as the keyword “atypical pneumonia” in the corresponding CT reports. In order to exclude undiagnosed (RT-PCR-negative)

COVID-19 cases with secondary pulmonary infections, we restricted this search to the time period from 01/2017 to 01/2019. This yielded a total of 196 CT exams. All CT scans were reviewed by two second-year radiology residents, who did not participate in the following reader study. CT scans were included if significant pulmonary findings, relating to pneumonia, were apparent. The associated viral, bacterial, and fungal pathogens were determined from the microbiological test results of oral swabs, bronchial lavage, tissue samples, and/or positive blood tests for specific pathogens. Comorbidities were assessed using the Charlson Comorbidity Index (CCI), which predicts the one-year mortality for a patient who may have a range of comorbid conditions, such as heart disease or cancer [22]. Both cohorts were matched for age, sex, comorbidities, as reflected by the CCI and type of inpatient treatment (emergency department/normal ward vs. intensive care unit) as well as necessity of invasive ventilation. 60 patients were included in this group.

Image Acquisition

All CT studies had been acquired on two CT systems (Somatom Definition AS-40 and Definition FLASH, Siemens Medical Systems, Forchheim, Germany). Depending on the clinical situation of the patient and the differential diagnoses considered at the time of the examination, standardized chest CT examinations were conducted using either a low- or full-dose technique, with a tube voltage of 80 kV or 120–140 kV, respectively, with the tube current automatically modulated (CareDose4D). CT images were reconstructed with a 1-mm and 3-mm slice thickness.

Data analysis

All images were anonymized and stored in a local PACS folder (IntelliSpace PACS 4.4 Radiology, Philips Medical Systems, Best, The Netherlands) in random order. Studies were then interpreted independently by four individuals: Two radiology residents with 1 (A) and 2 (B) years of experience, and by two board-certified radiologists with 7 (C) and 15 (D) years of experience, respectively.

All four radiologists had been trained in the radiological assessment of COVID-19 pneumonia by published as well as in-house case studies. As published in previous papers, the criteria for image evaluation were distribution patterns of lesions such as ground-glass opacities, consolidations, interlobular and interlobar thickening and crazy-paving pattern [2, 3, 23–25].

Findings in each CT study had to be categorized according to the likelihood with which COVID-19-associated lung disease was present using COV-RADS (Corona Virus imaging Reporting and Data System) [26]:

- COV-RADS 1: no pathological findings
- COV-RADS 2: CT findings suggestive of pneumonia or other lung disease without evidence of COVID-19
- COV-RADS 3: findings that could represent COVID-19-associated pneumonia
- COV-RADS 4: findings suspicious of COVID-19-associated pneumonia
- COV-RADS 5: findings typical of COVID-19-associated pneumonia

The respective patients' RT-PCR or microbiological findings were used as the standard of reference. Diagnoses categorized as COV-RADS 3–5 in patients with RT-PCR positive for SARS-CoV-2 were considered true-positive. Diagnoses of COV-RADS 1–2 in patients found to have no or non-COVID-19-associated lung disease were considered true-negative. Diagnoses categorized as COV-RADS 1–2 in patients with RT-PCR positive for SARS-CoV-2 were considered false-negative. Diagnoses of COV-RADS 3–5 in patients without evidence of SARS-CoV-2 infection were considered false-positive.

In addition to COV-RADS scores, the patterns and distribution of the respective pneumonias were objectively quantified in a blinded fashion in all 120 CT examinations. Details are provided in ► **Table 1**.

The diagnostic categories assigned by radiologists, the imaging findings as well as the respective patient's demographics, comorbidities and type of pathogen were collected in a pseudonymized database and statistically analyzed using IBM SPSS 26 (Armonk, NY, USA). Patient demographics, comorbidities, and imaging findings were compared using mean, standard deviation, Mann-Whitney-U-Tests, Chi²-Tests and Fischer-Exact-Tests. Diagnostic accuracy as well as Youden's Index were calculated using contingency tables. 95 % confidence intervals were calculated using Clopper-Pearson tests. Reviewer concordance was assessed using Fleiss' kappa tests.

Results

Patient's demographics and comorbidities

The final test cohort consisted of 120 patients who suffered either from COVID-19 (n = 60) or other atypical pneumonias (n = 60). Patient age and sex distribution are summarized in ► **Table 2**. Age and sex distribution were not significantly different between both groups (p = 0.17 and p = 0.17). Comorbidities, as reflected by the CCI, were slightly lower in the COVID-19 group with a mean of 3.75 (± 2.50) compared to the non-COVID-19 group (4.87 ± 2.85; p = 0.011).

CT imaging of patients with COVID-19 was performed on the same day as the positive RT-PCR test in 56/60 patients, with a mean delay of 0.14 days. At the time of the study CT examination, 60 % (36/60) of the patients from the COVID-19-positive group were either temporary patients of the emergency department or inpatients in standard care wards, and 40 % (24/60) were receiving treatment in intensive care units, with 15/60 (25 %) patients being treated with invasive ventilation. Patients from the non-COVID-19 group had viral infections in 19/60 cases (31 %), fungal infections in 16/60 cases (26 %), and bacterial infections in 26/60 cases (42 %). The detailed distribution of pathogens is provided in ► **Fig. 1**. At the time of the study CT examination, 53 % (32/60) patients from the non-COVID-19 group were inpatients in standard wards, and 47 % (28/60) were receiving treatment in intensive care units (ICUs), with 11/60 (18.3 %) patients being treated with invasive ventilation.

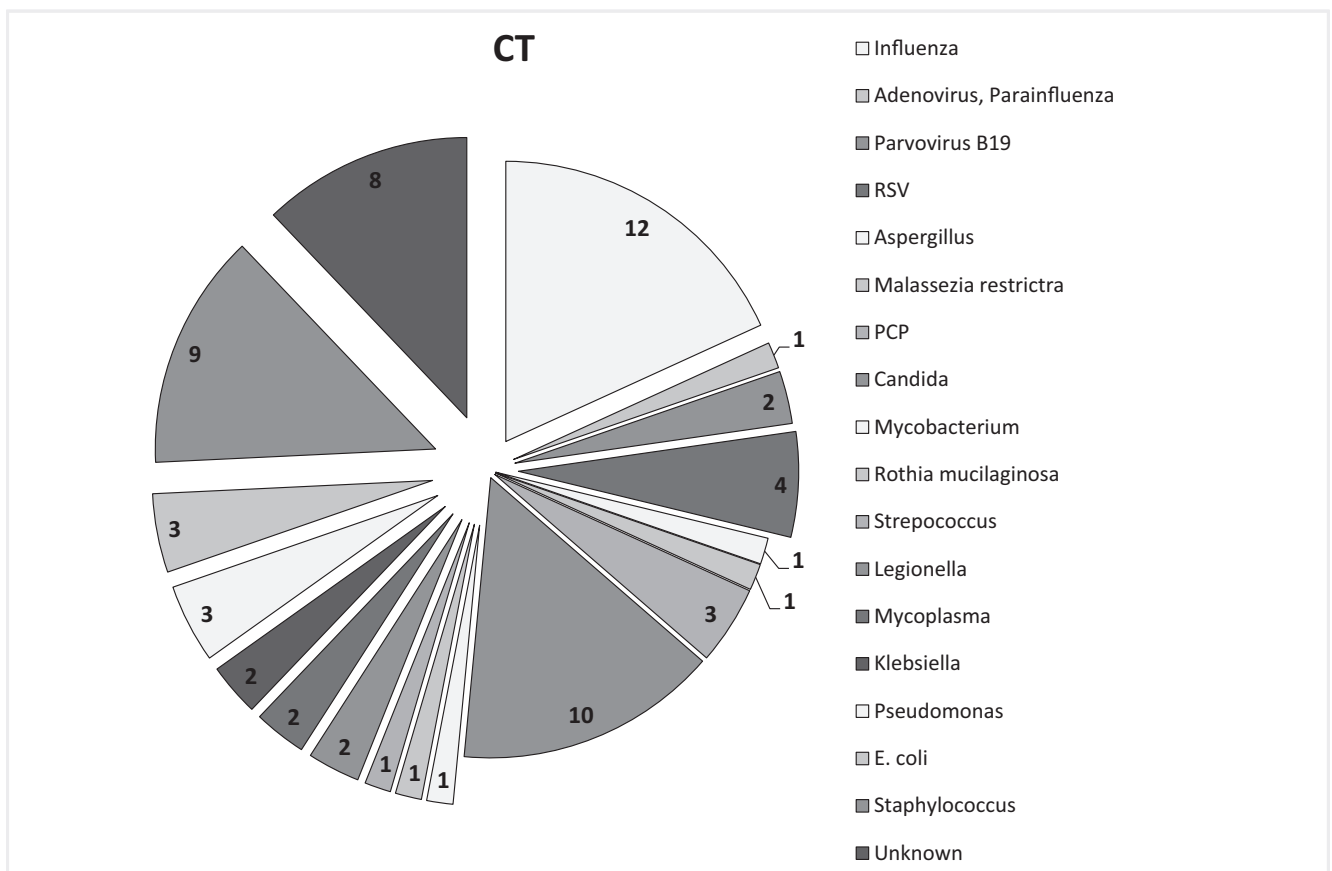
► **Table 1** Imaging features. These were evaluated in consensus by two second-year residents and a 7-year chest radiologist. The degree of involvement was subjectively evaluated.

► **Tab. 1** Bildmorphologische Veränderungen. Diese wurden durch 2 Assistenzärzte im zweiten Jahr und einen radiologischen Oberarzt mit 7 Jahren Berufserfahrung im Konsens eruiert. Der Beteiligungsgrad wurde subjektiv bewertet.

| imaging features | COVID-19 | % | COVID-19 mimics | % | p-value |
|-------------------------------|----------|---------|-----------------|---------|---------|
| ground-glass opacities (GGOs) | 59/60 | (98%) | 60/60 | (100%) | 0.99 |
| consolidations | 37/60 | (61.7%) | 41/60 | (68.3%) | 0.57 |
| GGOs and consolidations | 37/60 | (61.7%) | 41/60 | (68.3%) | 0.57 |
| Vertical distribution | | | | | 0.14 |
| apical emphasis | 9/59 | (15.3%) | 16/60 | (26.7%) | |
| basal emphasis | 10/59 | (16.9%) | 14/60 | (23.3%) | |
| no unambiguous emphasis | 40/59 | (67.8%) | 30 /60 | (50.0%) | |
| GGOs | | | | | |
| symmetry | | | | | 0.51 |
| unilateral | 6/59 | (10.2%) | 7/60 | (11.7%) | |
| bilateral | 53/59 | (89.8%) | 53/60 | (88.3%) | |
| axial distribution pattern | | | | | 0.52 |
| centrally emphasized | 0/59 | (0.0%) | 12/60 | (20%) | |
| peripherally emphasized | 37/59 | (62.7%) | 13/60 | (21.7%) | |
| no unambiguous emphasis | 22/59 | (37.3%) | 35/60 | (58.3%) | |
| degree of involvement | | | | | 0.35 |
| 0: | 1/60 | (1.7%) | 0/60 | (0.0%) | |
| 1: | 8/60 | (13.3%) | 12/60 | (20.0%) | |
| 2: | 22/60 | (36.7%) | 11/60 | (18.3%) | |
| 3: | 12/60 | (20.0%) | 14/60 | (23.3%) | |
| 4: | 8/60 | (13.3%) | 12/60 | (20.0%) | |
| 5: | 9/60 | (15.0%) | 11/60 | (18.3%) | |
| consolidations | | | | | |
| symmetry | | | | | 0.99 |
| unilateral | 9/37 | (24.3%) | 9/41 | (22.0%) | |
| bilateral | 28/37 | (75.7%) | 32/41 | (78.0%) | |
| axial distribution pattern | | | | | 0.15 |
| centrally emphasized | 1/37 | (2.7%) | 4/41 | (9.8%) | |
| peripherally emphasized | 30/37 | (81.1%) | 25/41 | (61.0%) | |
| no unambiguous emphasis | 6/37 | (16.2%) | 12/41 | (29.3%) | |
| degree of involvement | | | | | 0.28 |
| 0: | 23/60 | (38.3%) | 19/60 | (31.7%) | |
| 1: | 13/60 | (21.7%) | 9/60 | (15.0%) | |
| 2: | 10/60 | (16.7%) | 17/60 | (28.3%) | |
| 3: | 9/60 | (15.0%) | 7/60 | (11.7%) | |
| 4: | 4/60 | (6.7%) | 4/60 | (6.7%) | |
| 5: | 1/60 | (1.7%) | 4/60 | (6.7%) | |

► **Table 1** (Continuation)

| imaging features | COVID-19 | % | COVID-19 mimics | % | p-value |
|------------------------------------------|----------|----------|-----------------|----------|---------|
| other signs of atypical pneumonia | | | | | |
| crazy-paving pattern | 16/60 | (26.7 %) | 7/60 | (11.7 %) | 0.06 |
| inverse halo sign | 1/60 | (1.7 %) | 3/60 | (5.0 %) | 0.62 |
| halo sign | 1/60 | (1.7 %) | 1/60 | (1.7 %) | 0.99 |
| cavities | 2/60 | (3.3 %) | 1/60 | (1.7 %) | 0.99 |
| tree-in-bud pattern | 2/60 | (3.3 %) | 8/60 | (13.3 %) | 0.095 |
| nodular pattern | 0/60 | (0.0 %) | 4/60 | (6.7 %) | 0.12 |
| pleural effusion | 9/60 | (15.0 %) | 24/60 | (40.0 %) | 0.004 |



► **Fig. 1** Pathogen distribution of patients with COVID-19 negative, atypical pneumonia (COVID-19 mimics).

► **Abb. 1** Erregerverteilung der Kontrollgruppe mit Nicht-SARS-COV-2-assoziiertes, atypischer Pneumonie (COVID-19 mimics).

Results of reader study

A total of 480 CT ratings were collected from the four readers, 240 ratings from CT scans of patients suffering from COVID-19-associated pneumonia and 240 ratings from patients with non-COVID-19-associated pneumonia (“COVID-19 mimics”). Results for each reader (A–D), including contingency tables, are provided in ► **Table 3**.

Pooled among the four readers, CT ratings were true-positive for COVID-19 (COVID-19 present and ratings categorized as COV-RADS score 3–5) in 219/240 (91.3 %) cases, and true-negative (non-COVID-19-associated pneumonia present and COV-RADS score 1–2) in 159/240 (66.3 %) cases. This yields an overall diagnostic accuracy for the differentiation of COVID-19-associated pneumonia from non-COVID-19-associated atypical pneumonia of 78.8 % (378/480). A total of 21/240 (8.8 %) cases were clas-

► **Table 2** Patient demographics and comorbidities as well as stationary treatment and necessity for invasive ventilation for patients with COVID-19-associated pneumonia and non-COVID-19-associated, other atypical pneumonia ("COVID-19 mimics").

► **Tab. 2** Patientendemografie und Begleiterkrankungen sowie Führungsstatus (Normalstation/Intensivstation) und Beatmungspflicht der COVID-19- und non-COVID-19-Kohorte (andere atypische Pneumonien; „Mimics“).

| patient demographics | COVID-19 | COVID-19 mimics | p-value |
|---------------------------------------------------|---------------------|----------------------|---------|
| age | 66.1 (± 12.4) | 62.4 (± 17.2) | 0.168 |
| gender | 38/60 male (63.3 %) | 45/60 male (75 %) | 0.166 |
| charlson comorbidity index | 3.75 (± 2.50) | 4.87 (± 2.85) | 0.011 |
| emergency department or standard ward | 36/60 (60 %) | 32/60 (53 %) | 0.461 |
| intensive care unit | 24/60 (40 %) | 28/60 (47 %) | 0.461 |
| invasive ventilation at time of image acquisition | 15/60 (25 %) | 11/60 (18.3 %) | 0.506 |

► **Table 3** Detailed results of the reader study with contingency tables and test quality criteria. True positive was defined as a CT diagnosis positive for COVID-19 (COV-RADS 3, 4 or 5) in patients who were confirmed to have COVID-19 based on RT-PCR. True negative was defined as a CT diagnosis negative for COVID-19 (COV-RADS 1 or 2) in a patient with atypical pneumonia due to pathogens other than SARS-COV-2, i. e., other viral, fungal, or bacterial agents.

► **Tab. 3** Detaillierte Ergebnisse der Reader-study mit 4-Feldertafeln und Testgütekriterien. Richtig positive („True positive“) für COVID-19 wurden definiert als COV-RADS 3, 4 oder 5 mit IRT-PCR bestätigter COVID-19-Infektion. Richtig negative („True negative“) für COVID-19 wurden definiert als COV-RADS 1 oder 2 bei Patienten mit anderer, nicht SARS-COV-2-assoziiertes atypischer Pneumonie (z. B. viral, Pilz- oder Bakterienpneumonie).

| reader | results | | | | | | |
|---------------------------------------------|--------------|------|------|-----|---------------------------------------|-----------------------|--------------------|
| | | CT + | CT – | sum | diagnostic accuracy | test quality criteria | |
| A first-year resident | SARS-COV-2 + | 58 | 2 | 60 | 89/120 (74.4 %) [65.4 % – 81.7 %] | sensitivity | 96.7 % [88.5–99.6] |
| | SARS-COV-2 – | 29 | 31 | 60 | | specificity | 51.7 % [38.4–64.8] |
| | sum | 87 | 33 | 120 | | PPV | 66.7 % [55.7–76.4] |
| | | | | | | NPV | 93.9 % [79.8–99.3] |
| B second-year resident | SARS-COV-2 + | 53 | 7 | 60 | 94/120 (78.3 %) [69.9 % – 85.3 %] | sensitivity | 88.3 % [77.4–95.2] |
| | SARS-COV-2 – | 19 | 41 | 60 | | specificity | 68.3 % [55.0–79.7] |
| | sum | 72 | 48 | 120 | | PPV | 73.6 % [61.9–83.3] |
| | | | | | | NPV | 85.4 % [72.2–93.9] |
| C radiologist 7 years | SARS-COV-2 + | 56 | 4 | 60 | 101/120 (84.2 %) [76.4 % – 90.2 %] | sensitivity | 93.3 % [83.8–98.2] |
| | SARS-COV-2 – | 15 | 45 | 60 | | specificity | 75.0 % [62.1–85.3] |
| | sum | 71 | 49 | 120 | | PPV | 78.9 % [67.6–87.7] |
| | | | | | | NPV | 91.8 % [83.4–97.7] |
| D chest radiolo- gist 15 years | SARS-COV-2 + | 52 | 8 | 60 | 94/120 (78.3 %) [69.9 % – 85.3 %] | sensitivity | 86.7 % [75.4–94.1] |
| | SARS-COV-2 – | 18 | 42 | 60 | | specificity | 70.0 % [56.8–81.2] |
| | sum | 70 | 50 | 120 | | PPV | 74.3 % [62.4–84.0] |
| | | | | | | NPV | 84.0 % [70.9–92.8] |
| A+B+C+D | | CT + | CT – | sum | diagnostic accuracy | test quality criteria | |
| | SARS-COV-2 + | 219 | 21 | 240 | 378/480 (78.8 %) [74.8 % – 82.3 %] | sensitivity | 91.3 % [86.9–94.5] |
| | SARS-COV-2 – | 81 | 159 | 240 | | specificity | 66.3 % [59.9–72.2] |
| | sum | 300 | 180 | 480 | | PPV | 73.4 % [67.6–77.9] |
| | | | | NPV | | 88.8 % [82.7–92.6] | |



► **Fig. 2** Influenza-associated pneumonia. Extensive ground-glass opacities and consolidation in the left lower lobe. Focal ground-glass opacities in the left upper lobe and right lower lobe, both partly forming a tree-in-bud pattern. Bilateral pleural effusion. All four radiologists rated this exam as COV-RADS 2.

► **Abb. 2** Influenza-assoziierte Pneumonie. Ausgedehnte Milchglastrübungen und Konsolidierungen im linken Unterlappen. Teils fokale Milchglastrübungen im rechten Unterlappen und linken Oberlappen, teils mit teilerfasstem Tree-in-bud-Muster. Begleitender, bilateraler Pleuraerguss. Von allen 4 Radiologen als COV-RADS 2 bewertet.



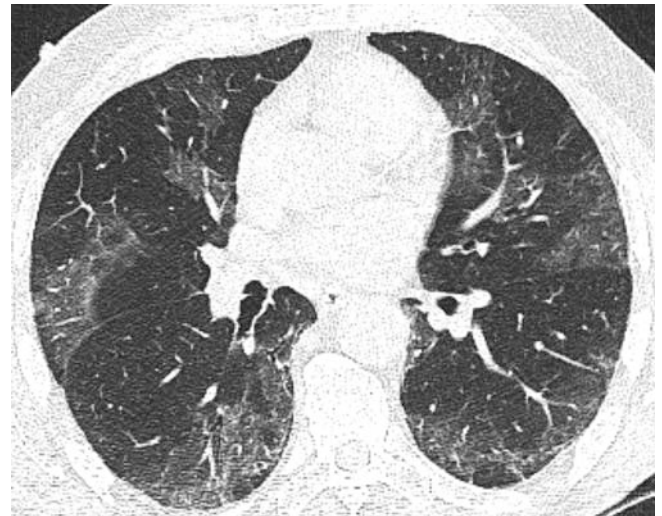
► **Fig. 4** COVID-19-associated pneumonia. Patchy ground-glass opacities, mostly located peripherally in this slice particularly visible in both upper lobes. All four radiologists rated this exam as COV-RADS 4.

► **Abb. 4** COVID-19-assoziierte Pneumonie. Fleckige, peripher betonte Milchglastrübungen, hier vor allem in beiden Oberlappen. Von allen 4 Radiologen als COV-RADS 4 bewertet.



► **Fig. 3** COVID-19-associated pneumonia. Scattered, ribbon-shaped consolidations in all lung lobes with surrounding ground-glass opacities, mostly located peripherally; in this slice particularly visible in the right upper lobe. All four radiologists rated this exam as COV-RADS 3.

► **Abb. 3** COVID-19-assoziierte Pneumonie. Einzelne, peripher betonte, bandförmige Konsolidierungen mit umgebender Milchglastrübung, in allen Lungenlappen mit peripherer Betonung; hier vor allem im rechten Oberlappen. Von allen 4 Radiologen als COV-RADS 3 bewertet.

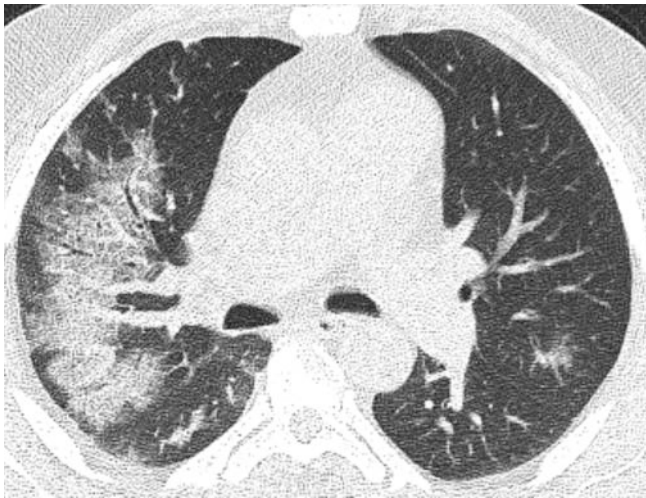


► **Fig. 5** COVID-19-associated pneumonia. Multiple, mostly peripherally distributed ground-glass opacities in all lung lobes. All four radiologists rated this exam as COV-RADS 5.

► **Abb. 5** COVID-19-assoziierte Pneumonie. Multiple, peripher betonte Milchglastrübungen in allen Lungenlappen. Von allen 4 Radiologen als COV-RADS 5 bewertet.

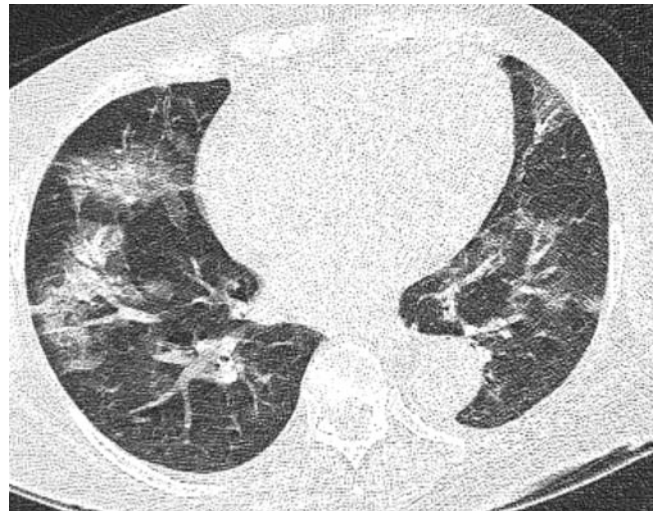
sified as false-negative (patients with confirmed COVID-19-associated pneumonia categorized as COV-RADS 1–2). The remaining

81/240 (33.8%) were classified as false-positive (patients with non-COVID-19-associated atypical pneumonia categorized as COV-RADS 3–5). Pooled Youden's Index was 0.57. Exemplary images of a patients rated with COV-RADS 2–5 are demonstrated in ► **Fig. 2–7**.



► **Fig. 6** Candida pneumonia with low inter-reader concordance (pitfall case). Multiple ground-glass opacities in all lung lobes with an emphasis on the right side (partially with crazy-paving pattern). This was accompanied by consolidations. Two radiologists rated this exam as COV-RADS 5, one as COV-RADS 3 and one as COV-RADS 2.

► **Abb. 6** Candida-Pneumonie mit geringer Bewerter-Übereinstimmung (Pitfall-Fall). Multiple Milchglastrübungen in allen Lungenlappen, rechtsseitig betont (teils mit Crazy-paving-Muster). Begleitende Konsolidierungen. Von 2 Radiologen als COV-RADS 5, von einem als COV-RADS 3 und einem als COV-RADS 2 bewertet.



► **Fig. 7** Influenza-associated pneumonia with low case-specific diagnostic accuracy (pitfall case). Multiple ground-glass opacities in all lung lobes without an unambiguous emphasis (partially with crazy-paving pattern). This was accompanied by minor consolidations, as well as bronchiectasis. One radiologist rated this exam as COV-RADS 5, two as COV-RADS 4 and one as COV-RADS 3.

► **Abb. 7** Influenza-assoziierte Pneumonie mit geringer fallspezifischer diagnostischer Genauigkeit (Pitfall-Fall). Multiple Milchglastrübungen ohne eindeutigen Verteilungsgradienten in allen Lungenlappen. Begleitend diskrete Konsolidierungen und Bronchiektasen. Von einem Radiologen als COV-RADS 5, von 2 als COV-RADS 4 und einem als COV-RADS 3 bewertet.

A subgroup analysis of patients managed in normal wards and the emergency department versus intensive care patients indicated that specificity in more severe pneumonia was overall lower. Yet while confidence intervals overlapped, no significant difference was observed neither for sensitivity (93.8% [95%-CI: 88.5–97.1%] and 87.5% [95%-CI: 79.2–93.4%], $p=0.11$) nor specificity (70.3% [95%-CI: 61.6–78.1%] and 62.5% [95%-CI: 52.9–71.5%], $p=0.22$). Reviewers A and B were radiology residents with 1 and 2 years of experience, respectively. Reviewers C and D were chest radiologists with 7 and 15 years of experience, respectively. Youden's Indices for each individual reviewer were 0.48 (A), 0.57 (B), 0.68 (C), and 0.57 (D), respectively. The individual diagnostic accuracies observed for the four readers did not differ to a statistically significant degree ($p=0.3$). Between-reader agreement was moderate with $\kappa=0.53$. Details are provided in ► **Table 4**.

Pulmonary findings

All 120 CT scans were analyzed for the pattern and distribution of pneumonia. Detailed results are shown in ► **Table 1**. In summary, both cohorts showed similar expressions of pulmonary GGOs and consolidations ($p=0.99$ and 0.57 , respectively) as well as a comparable vertical distribution of patterns ($p=0.14$). There was a strong tendency towards a higher occurrence of crazy-paving in the COVID-19 group (27% vs. 12%, $p=0.06$), whereas the tree-in-bud pattern and nodular patterns tended to be more frequent in the COVID-19 mimics (3% vs. 13% and 0% vs. 7%, $p=0.10$ and 0.12 , respectively). Pleural effusion was more common in the COVID mimic cohort, which was statistically significant (15% vs. 40%, $p=0.004$).

Summary of main results

Overall, the reviewers demonstrated a sensitivity of 91% (95%-CI: 87–95%) and a specificity of 66% (95%-CI: 60–72%). Diagnostic accuracy was 79% (95%-CI: 75–82%). Individual results did not differ significantly, while between-reader concordance was moderate.

Discussion

This study evaluates the diagnostic accuracy with which COVID-19-associated pneumonia is distinguishable from other atypical pneumonias in chest CT, and whether reader expertise is a driver of the observed accuracy.

It is important to emphasize that in this context, sensitivity and specificity do not refer to the ability to tell diseased from healthy subjects – but to the ability to distinguish different types of atypical pneumonia in diseased patients, carefully selected to exhibit similar demographic features, clinical situation, fraction of patients under ventilation, and general medical condition as indicated by a similar CCI. We deliberately constructed the most difficult setting in which to investigate the diagnostic accuracy of COVID-19 pneumonia on CT, without a control group comprised of healthy patients. It was therefore almost to be expected that the specificity would be lower than in studies with “real life” patient cohorts. Nevertheless, we found that all four readers were able to correctly identify the presence of COVID-19-associated pneumonia with a sensitivity of 91%

► **Table 4** Subgroup analysis with test quality criteria and Clopper-Pearson 95 % confidence intervals as well as Fleiss' kappa.

► **Tab. 4** Subgruppenanalysen mit Testgütekriterien und Clopper-Pearson- 95 %-Konfidenzintervallen sowie Fleiss' Kappa.

| total test quality criteria | | | | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|------------------------------------|---------------------|----------------------------|-----------------------------|
| reader | true positive rate* (sensitivity) | true negative rate** (specificity) | diagnostic accuracy | positive predictive value* | negative predictive value** |
| A | 96.7 % [88.5–99.6] | 51.7 % [38.4–64.8] | 74.2 % [65.4–81.7] | 66.7 % [55.7–76.4] | 93.9 % [79.8–99.3] |
| B | 88.3 % [77.4–95.2] | 68.3 % [55.0–79.7] | 78.3 % [69.9–85.3] | 73.6 % [61.9–83.3] | 85.4 % [72.2–93.9] |
| C | 93.3 % [83.8–98.2] | 75.0 % [62.1–85.3] | 84.2 % [76.4–90.2] | 78.9 % [67.6–87.7] | 91.8 % [83.4–97.7] |
| D | 86.7 % [75.4–94.1] | 70.0 % [56.8–81.2] | 78.3 % [69.9–85.3] | 74.3 % [62.4–84.0] | 84.0 % [70.9–92.8] |
| mean | 91.3 % [86.9–94.5] | 66.3 % [59.9–72.2] | 78.8 % [74.8–82.3] | 73.4 % [67.6–77.9] | 88.8 % [82.7–92.6] |
| κ = 0.53 | | | | | |
| subgroup analysis with ICU inpatients only | | | | | |
| reader | true positive rate* (sensitivity) | true negative rate** (specificity) | diagnostic accuracy | positive predictive value* | negative predictive value** |
| A | 91.7 % [73.0–99.0] | 50.0 % [30.6–69.4] | 69.2 % [54.9–81.3] | 61.1 % [43.5–76.9] | 87.5 % [61.7–98.4] |
| B | 83.3 % [62.6–95.3] | 71.4 % [51.3–86.8] | 76.9 % [63.2–87.5] | 71.4 % [51.3–86.8] | 83.3 % [62.6–95.3] |
| C | 95.8 % [78.9–100.0] | 71.4 % [51.3–86.8] | 82.7 % [69.7–91.8] | 74.2 % [55.4–88.1] | 95.2 % [76.2–100.0] |
| D | 79.2 % [57.8–92.9] | 57.1 % [37.2–75.5] | 67.3 % [52.8–79.7] | 61.3 % [42.2–78.2] | 76.2 % [52.8–91.8] |
| mean | 87.5 % [79.2–93.4] | 62.5 % [52.9–71.5] | 74.0 % [67.5–79.9] | 67.0 % [57.7–74.8] | 85.6 % [75.8–92.2] |
| κ = 0.53 | | | | | |
| subgroup analysis excluding ICU inpatients | | | | | |
| reader | true positive rate* (sensitivity) | true negative rate** (specificity) | diagnostic accuracy | positive predictive value* | negative predictive value** |
| A | 100.0 % [90.3–100.0] | 53.1 % [34.7–70.9] | 77.9 % [66.2–87.1] | 70.6 % [56.2–82.5] | 100.0 % [80.5–100.0] |
| B | 91.7 % [77.5–98.2] | 68.8 % [50.0–83.9] | 80.9 % [69.5–89.4] | 76.7 % [61.4–88.2] | 88.0 % [68.8–97.5] |
| C | 91.7 % [77.5–98.2] | 78.1 % [60.0–90.7] | 85.3 % [74.6–92.7] | 82.5 % [67.2–92.7] | 89.3 % [71.8–97.7] |
| D | 91.7 % [77.5–98.2] | 81.3 % [63.6–92.8] | 86.8 % [76.4–93.8] | 84.6 % [69.5–94.1] | 89.7 % [72.6–97.8] |
| mean | 93.8 % [88.5–97.1] | 70.3 % [61.6–78.1] | 82.7 % [77.7–87.0] | 78.6 % [71.1–84.0] | 91.7 % [83.4–95.8] |
| κ = 0.53 | | | | | |
| * True positive was defined as a CT diagnosis positive for COVID-19 (COV-RADS 3, 4 or 5) in patients who were confirmed to have COVID-19 based on RT-PCR. Richtig Positive („True positive“) für COVID-19 wurden definiert als COV-RADS 3, 4 oder 5 mit laborchemisch bestätigter COVID-19-Infektion. | | | | | |
| ** True negative was defined as a CT diagnosis negative for COVID-19 (COV-RADS 1 or 2) in a patient with atypical pneumonia due to pathogens other than SARS-COV-2, i. e., other viral, fungal, or bacterial agents Richtig Negative („True negative“) für COVID-19 wurden definiert als COV-RADS 1 oder 2 bei Patienten mit anderer, nicht SARS-COV-2-assoziierter atypischer Pneumonie (z. B. viral, Pilz- oder Bakterienpneumonie). | | | | | |

(95 %-CI: 87–95 %), and a specificity of 66 % (95 %-CI: 60–72 %). The overall diagnostic accuracy for the differentiation of COVID-19-associated pneumonia from non-COVID-19-associated atypical pneumonia was 79 % (95 %-CI: 75–82 %).

In contrast, the ACR still does not recommend chest CT for the diagnosis of COVID-19 pneumonia since the specificity is considered too low. Indeed, one of the first studies on the systematic use of chest CT for the diagnosis of COVID-19 reported a specificity of 25 % [3]. However, it should be noted that, in the early stage of the current pandemic, radiological signs of COVID-19 were not yet fully understood, and the reading behavior of radiologists was presumably different than today. Additionally, PCR tests pre-

sumably were not as reliable as they are today. Therefore, a considerable number of false-negative results must also be assumed. A recent meta-analysis of 31 CT studies and 8014 patients reported a pooled sensitivity and specificity of 89.9 % and 61.1 %, respectively. However, it is important to note that the included studies were in part rather heterogeneous with specificities ranging from 0–96 %. 53 % of the studies were from China and 26 % were still in preprint status at the time of the publication of this meta-analysis [19].

On the contrary, a recent multicenter study from France reported sensitivity and specificity rates of 90 % (95 %-CI: 89–91)

and 91 % (95 %-CI: 91–92 %), respectively, in a cohort of over 4800 patients [14].

Our results indicate that, despite these difficulties, specificity was at least moderate even in this direct discrimination between atypical pneumonias. This is also in line with other recent European publications [26–28].

The individual diagnostic accuracies varied only mildly between readers and ranged from 74 % to 84 %. While the respective sensitivity levels were relatively stable across readers, ranging from 88 % to 97 %, the observed specificity levels were less consistent, ranging from 52 % to 75 %. Although we had included readers with different professional backgrounds, ranging from a first-year resident to a dedicated thoracic radiologist with 15 years of experience, neither overall accuracy nor sensitivity or specificity correlated with reader expertise. This indicates that the evaluation of COVID-19 is similarly challenging for both experienced radiologists and junior residents, most likely due to the recency of the onset of the disease. The assessment of CT examinations was performed using the COV-RADS classification, which predicts the likelihood of COVID-19 pneumonia. Since these ratings are subjective reader decisions, pulmonary findings cannot be derived and were additionally determined for better comparability (► **Table 1**). We found comparable extents of GGOs and consolidations in both cohorts, with fairly similar distributions. The subgroup analysis of patients in normal wards and intensive care units indicated that specificity in severe courses of pneumonia was lower, though this was not statistically significant.

Our study has several limitations. First, this is a study on a small patient cohort. Second, we explicitly withheld any information on clinical symptoms and/or results of laboratory tests to readers. In the clinical routine, such information is usually available when interpreting CT imaging data. Consequently, readers were not able to assess which COVID-19-related changes to monitor based on previously reported image evolution characteristics influenced, for example, by time since symptom onset [29, 30]. It is unclear whether such data would further increase or decrease the diagnostic accuracy of chest CT for the differentiation of COVID-19 pneumonia versus non-COVID-19-associated pneumonia. Third, regarding the lack of influence of reader expertise on diagnostic accuracy, it has to be kept in mind that COVID-19 is a new disease that has been observed for only a couple of months. Accordingly, the “reader expertise” of the four readers is of course similar in this regard, and mostly differs for the interpretation of non-COVID-19-associated disease. Fourth, as mentioned above, due to the artificial composition of the study cohort, the observed statistical test performance measures cannot and should not be directly translated to those of “real-life scenarios”.

In conclusion, in this enriched cohort exclusively consisting of patients suffering from atypical pneumonia, radiologists were able to distinguish COVID-19 pneumonia from other causes of atypical pneumonia with at least moderate diagnostic accuracy, regardless of the years of training of the radiologists.

CLINICAL RELEVANCE

- Based on contemporary literature largely consistent with our results, CT seems to be a valuable tool in the diagnostic process of suspected COVID-19 cases.
- In this considerably more challenging, artificial scenario in the absence of healthy cases, radiologists were able to distinguish between COVID-19 and other atypical pneumonias with moderate diagnostic accuracy. The extent of pulmonary findings complicates the differentiation between COVID-19 pneumonia and other atypical pneumonias.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- Huang C, Wang Y, Li X et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395: 497–506. doi:10.1016/s0140-6736(20)30183-5
- Bernheim A, Mei X, Huang M et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology* 2020. doi:10.1148/radiol.2020200642
- Ai T, Yang Z, Hou H et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* 2020. doi:10.1148/radiol.2020200642
- Cheng Z, Lu Y, Cao Q et al. Clinical Features and Chest CT Manifestations of Coronavirus Disease 2019 (COVID-19) in a Single-Center Study in Shanghai, China. *Am J Roentgenol* 2020: 1–6. doi:10.2214/ajr.20.22959
- Salehi S, Abedi A, Balakrishnan S et al. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. *Am J Roentgenol* 2020: 1–7. doi:10.2214/ajr.20.23034
- Han R, Huang L, Jiang H et al. Early Clinical and CT Manifestations of Coronavirus Disease 2019 (COVID-19) Pneumonia. *Am J Roentgenol* 2020: 1–6. doi:10.2214/ajr.20.22961
- Ye Z, Zhang Y, Wang Y et al. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *Eur Radiol* 2020. doi:10.1007/s00330-020-06801-0
- Zhou S, Wang Y, Zhu T et al. CT Features of Coronavirus Disease 2019 (COVID-19) Pneumonia in 62 Patients in Wuhan, China. *Am J Roentgenol* 2020: 1–8. doi:10.2214/ajr.20.22975
- Wang K, Kang S, Tian R et al. Imaging manifestations and diagnostic value of chest CT of coronavirus disease 2019 (COVID-19) in the Xiaogan area. *Clin Radiol* 2020. doi:10.1016/j.crad.2020.03.004
- Carotti M, Salaffi F, Sarzi-Puttini P et al. Chest CT features of coronavirus disease 2019 (COVID-19) pneumonia: key points for radiologists. *Radiol Med* 2020. doi:10.1007/s11547-020-01237-4
- Sun Z, Zhang N, Li Y et al. A systematic review of chest imaging findings in COVID-19. *Quant Imaging Med Surg* 2020; 10: 1058–1079. doi:10.21037/qims-20-564
- Li L, Qin L, Xu Z et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. *Radiology* 2020. doi:10.1148/radiol.2020200905
- Fang Y, Zhang H, Xie J et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 2020. doi:10.1148/radiol.2020200432

- [14] Herpe G, Lederlin M, Naudin M et al. Efficacy of Chest CT for COVID-19 Pneumonia in France. *Radiology* 2020. doi:10.1148/radiol.2020202568
- [15] American College of Radiology. ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. In 2020 May 11.
- [16] Hosseiny M, Kooraki S, Gholamrezanezhad A et al. Radiology Perspective of Coronavirus Disease 2019 (COVID-19): Lessons From Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. *Am J Roentgenol* 2020; 214: 1078–1082. doi:10.2214/ajr.20.22969
- [17] Franquet T. Imaging of pulmonary viral pneumonia. *Radiology* 2011; 260: 18–39. doi:10.1148/radiol.11092149
- [18] Ellis SJ, Cleverley JR, Muller NL. Drug-induced lung disease: high-resolution CT findings. *Am J Roentgenol* 2000; 175: 1019–1024. doi:10.2214/ajr.175.4.1751019
- [19] Islam N, Salameh JP, Leeflang MM et al. Thoracic imaging tests for the diagnosis of COVID-19. *Cochrane Database Syst Rev* 2020; 11: Cd013639. doi:10.1002/14651858.CD013639.pub3
- [20] Bai HX, Hsieh B, Xiong Z et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* 2020. doi:10.1148/radiol.2020200823
- [21] Luo N, Zhang H, Zhou Y et al. Utility of chest CT in diagnosis of COVID-19 pneumonia. *Diagn Interv Radiol* 2020. doi:10.5152/dir.2020.20144
- [22] Charlson ME, Pompei P, Ales KL et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40: 373–383. doi:10.1016/0021-9681(87)90171-8
- [23] Chung M, Bernheim A, Mei X et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology* 2020; 295: 202–207. doi:10.1148/radiol.2020200230
- [24] Song F, Shi N, Shan F et al. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology* 2020; 295: 210–217. doi:10.1148/radiol.2020200274
- [25] Pan F, Ye T, Sun P et al. Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia. *Radiology* 2020. doi:10.1148/radiol.2020200370
- [26] Schulze-Hagen M, Hübel C, Meier-Schroers M et al. Low-dose chest CT for the diagnosis of COVID-19 – a systematic, prospective comparison with PCR. *Dtsch Arztebl Int* 2020; 117: 389–395. doi:10.3238/arztebl.2020.0389
- [27] Gezer NS, Ergan B, Bariş MM et al. COVID-19 S: A new proposal for diagnosis and structured reporting of COVID-19 on computed tomography imaging. *Diagn Interv Radiol* 2020; 26: 315–322. doi:10.5152/dir.2020.20351
- [28] Schalekamp S, Bleeker-Rovers CP, Beenen LFM et al. Chest CT in the Emergency Department for Diagnosis of COVID-19 Pneumonia: Dutch Experience. *Radiology* 2020. doi:10.1148/radiol.2020203465
- [29] Kovács A, Palásti P, Veréb D et al. The sensitivity and specificity of chest CT in the diagnosis of COVID-19. *Eur Radiol* 2020: 1–6. doi:10.1007/s00330-020-07347-x
- [30] Wang Y, Dong C, Hu Y et al. Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. *Radiology* 2020; 296: e55–e64. doi:10.1148/radiol.2020200843