La radiologia medica (2021) 126:577–584
https://doi.org/10.1007/s11547-020-01320-w

CHEST RADIOLOGY

Inter-reader agreement of high-resolution computed tomography findings in patients with COVID-19 pneumonia: A multi-reader study

Lorenzo Cereser[1] · Rossano Girometti[1] · Jacopo Da Re[1] · Filippo Marchesini[1] · Giuseppe Como[1] · Chiara Zuiani[1]

Received: 23 July 2020 / Accepted: 26 November 2020 / Published online: 3 January 2021

Abstract

Purpose To investigate the inter-reader agreement in assessing high-resolution computed tomography (HRCT) features of coronavirus disease 2019 (COVID-19) pneumonia.

Method Seventy-seven consecutive patients (mean age, 64 ± 15 years) with mild COVID-19 pneumonia that underwent HRCT were retrospectively included. Three radiologists [two devoted to thoracic imaging (R1, R2), and one generalist (R3)] on a per-examination basis independently assessed ground-glass opacity (GGO), consolidation, and crazy-paving pattern. The extent of each feature (total feature score, TFS) was semi-quantitatively assessed, and each TFS summed up to obtain total lung score (TLS). Presence of organizing pneumonia (OP) pattern was also recorded. The inter-reader agreement was calculated with Cohen's Kappa (k) and Free-Marginal Multirater k. Multivariable analysis was run to determine whether imaging features were predictive of short-term evolution to severe disease (need for ventilation).

Results Most features showed substantial inter-reader agreement, including TLS > 6 ($k = 0.69$), which was an independent predictor of short-term occurrence of severe disease, regardless of the reader (OR 9–53.19). Consolidation TFS > 2 and OP pattern showed substantial and moderate agreement, respectively, only when comparing R1 and R2. Consolidation TFS > 2 and OP pattern were independent predictors of severe disease for R2 (OR 4.87) and R1 (OR 6), respectively.

Conclusions The inter-reader agreement for most HRCT features of COVID-19 pneumonia ranges moderate-to-substantial, though it depends on readers' experience in the case of consolidation and OP pattern.

Keywords High-resolution computed tomography · Pneumonia · Coronavirus · Inter-reader agreement

Introduction

Lung disease is the main manifestation of the COVID-19 [1], with clinical presentation ranging from asymptomatic to fever, dry cough, fatigue, and dyspnea, up to respiratory failure in severe cases [2]. The standard of reference for diagnosis is the reverse transcriptase polymerase chain reaction (RT-PCR) test, using nasal-pharyngeal swabs or lower respiratory tract specimens [3].

Chest HRCT plays an important role in the detection of COVID-19 pneumonia, with reported sensitivity ranging from 61% [4] to 99% in a study performed in a setting with high disease prevalence [5]. Typical findings include GGO and consolidation, involving multiple lobes of both lungs [6], as well as OP pattern [7]. Despite the low specificity (about 25–33%) [8, 9], HRCT with typical appearance may be of help when there is diagnostic uncertainty in a patient with high pretest probability for the disease [6]. In this light, an expert consensus statement from the Radiological Society

✉ Lorenzo Cereser
lcereser@sirm.org

Rossano Girometti
rgirometti@sirm.org

Jacopo Da Re
jacopo.dare6@gmail.com

Filippo Marchesini
filippo.marchesini88@gmail.com

Giuseppe Como
giuseppe.como@asufc.sanita.fvg.it

Chiara Zuiani
chiara.zuiani@uniud.it

[1] Institute of Radiology, Department of Medicine, University of Udine, University Hospital "S. Maria della Misericordia", p.le S. Maria della Misericordia, 15, 33100 Udine, Italy

of North America (RSNA) provided a system for categorizing HRCT findings based on the likelihood they represent COVID-19 pneumonia [7].

Another target of research is as to whether HRCT can predict unfavorable clinical outcome, which has been variably defined as progression to severe disease, Intensive Care Unit (ICU) admission, or death [10–13]. Previous Authors [10–13] found that qualitative and semi-quantitative indexes expressing the amount of lung involvement are associated to disease worsening.

One might assume that the pre-requisite for using HRCT as a diagnostic and predictive tool is adequate inter-reader agreement in assessing COVID-19 pneumonia-related HRCT features. To our knowledge, a few studies only investigated this topic [14–17]. Thus, it is uncertain whether interpretation and quantification of lung involvement can be reliably provided across different readers and different geographical areas involved by the pandemic [18]. Moreover, an adequate inter-reader agreement may further support the use of HRCT in many COVID-19 related clinical situations, e.g., in case of swab/clinical data doubts, in the evolution/worsening of the disease, and in the outcome evaluation.

The aim of the study was to investigate the inter-reader agreement in assessing HRCT features of COVID-19 pneumonia.

Material and methods

Patient population

The Ethical Committee approved the study protocol, and waived for the acquisition of the informed consent, given the retrospective design.

By performing a search in the database of our COVID-19-center, we identified all the consecutive adult patients with suspected COVID-19 pneumonia who underwent chest HRCT examination in the period March-April 2020. Before HRCT, all patients performed RT-PCR test for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in nasal-pharyngeal swabs, and were categorized according to the Italian Society of Emergency Medicine (SIMEU) classification of clinical phenotypes [19]. The latter included: (i) phenotype 1: fever without respiratory failure and normal chest X-ray; (ii) phenotype 2: fever with chest X-ray and arterial blood gas test indicating lung focus and/or mild respiratory failure [partial pressure of arterial blood oxygen (PaO2) > 60 mmHg]; (iii) phenotype 3: fever with moderate-severe respiratory failure (PaO2 < 60 mmHg in room air); (iv) phenotype 4: respiratory failure with suspected initial acute respiratory distress syndrome (ARDS) or complicated pneumonia; and (v) phenotype 5: overt ARDS [18]. Oxygen therapy and/or continuous positive airway pressure (CPAP) ventilation were indicated in patients with SIMEU phenotypes 3–4 pneumonia, while orotracheal intubation with invasive ventilation was the treatment for SIMEU phenotypes 4–5 [19].

Of the 192 eligible subjects, we excluded 104 patients with negative RT-PCR test, and 11 patients with clinical phenotypes 3–5 at the time of HRCT. Therefore, the final population consisted of 77 patients (40 men and 37 women, mean age 64 ± 15 years) with mild COVID-19 pneumonia (i.e., SIMEU clinical phenotypes 1–2). In cases the patient had undergone several HRCT examinations, only the baseline one was included in the analysis.

HRCT examinations

HRCTs were performed on a 64-row Computed Tomography (CT) scanner (LightSpeed, General Electric, Milwaukee, Wisconsin, USA), by means of volumetric acquisition with the patient in the supine position, at suspended full inspiration. Image acquisition parameters were as follows: 0.6 s gantry revolution time, 100–350 mA tube current modulation range, 120 kV tube potential, 64 mm × 0.625 mm detector configuration, 1.25 mm reconstructed section thickness and interval. In 4/77 patients (5.2%) iodinated contrast medium [iomeprol 350 mgI/mL (Iomeron, Bracco Imaging, Milan, Italy)] was intravenously injected before scanning. Two image sets were reconstructed and displayed, including one with high-spatial-frequency algorithm and pulmonary parenchyma windowing (level, −500 HU; width, 1700 HU), and the other with soft tissue algorithm and windowing (level, 50 HU; width, 350 HU).

Image analysis

For each patient, three readers recorded the presence of GGO, consolidation, and crazy-paving pattern, as defined by the glossary of terms for thoracic imaging from Fleischner Society [20]. Readers included two radiologists devoted to thoracic imaging, namely reader 1 (R1) and reader 2 (R2), with 10 and 3 years of experience, respectively, and one generalist radiologist (R3) with 20 years of experience in body imaging. Readers also assessed whether GGO and/or consolidation presented with an OP pattern, i.e., whether they showed triangular or polygonal shape, or were associated with perilobular pattern, bronchial dilatation, reverse halo sign, linear and band-like opacities, and signs of fibrosis [21]. On a per-examination basis, six lung zones were identified (3 per lung), i.e., two upper zones (above the carina), two middle zones (from the carina to the inferior pulmonary veins), and two lower zones (below the inferior pulmonary veins). Readers then assessed the zonal extent of GGO, consolidation, and crazy-paving pattern, using a previously reported semi-quantitative score [11, 22]. Score was 0 if

the feature was not present, 1 if it was present with a <25% zonal involvement, 2 for a ≥25% to <50% involvement, 3 for a ≥50% to <75% involvement, and 4 for ≥75% involvement. Therefore, the per-patient total score for each pulmonary feature (TFS) ranged from 0 (i.e., a certain feature was scored 0 in each of the six lung zones) to 24 (i.e., a certain feature was scored 4 in each of the six lung zones). TLS was defined as the summing up of the GGO, consolidation, and crazy-paving pattern TFSs.

Clinical data analysis

For all patients, comorbidities and time from symptoms onset to HRCT examination were reported. We recorded the patients' SIMEU phenotype twice, i.e., the one observed at the time of HRCT examination, and the worst one noticed in the 15-day period following HRCT. For the purpose of analysis, SIMEU phenotypes recorded during the follow-up period were dichotomized into mild disease group [including patients with no or mild respiratory failure (SIMEU phenotype 1 or 2)] versus severe disease group [including patients with moderate-to-severe respiratory failure or ARDS (SIMEU phenotype 3–5)].

On this basis, the study outcome was defined as the development of severe disease in the 15-day period following HRCT, i.e., a shift from SIMEU phenotype 1–2 to SIMEU phenotype 3–5.

Statistical analysis

We used descriptive statistics to summarize HRCT findings, and coupled relevant proportions with 95% confidence intervals (95%CI). After checking whether continuous data showed normal distribution, we described them as means ± standard deviation or medians with the interquartile range (IQR). The Cochran's Q test was used to determine whether there was any significant difference in the prevalence of each HRCT feature among the three sets of readings. Pairwise comparisons were performed with the McNemar test.

To determine the inter-reader agreement in assessing HRCT features we used Percent Agreement (PA), Cohen's Kappa (k), as recommended by the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [23], as well as Free-Marginal Multirater k. We used a TFS cut-off of >2, and a TLS cut-off >6 for including data into the analysis. When paradox k was observed [i.e., unacceptable kappa value ($k \leq 0.41$) and acceptable percent agreement (PA ≥ 0.80) with Prevalence Index and Bias Index different from zero], the imbalance was corrected by using the prevalence-adjusted bias-adjusted kappa (PABAK) statistic [24, 25]. Interpretation of k and PABAK coefficient was as follow: <0.00, poor; 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect [26].

On a per-reader basis, we then performed a logistic regression analysis with the stepwise approach to assess whether HRCT presentation could predict the occurrence of the study outcome as defined above. The model included TLS > 6, consolidation > 2, crazy-paving pattern > 2, and presence of OP pattern. Preliminary univariable analysis was performed with the chi-square test.

Analyses were performed using MedCalc statistical software (MedCalc Software bvba, version 18.11.6, Ostend, Belgium), and Online Kappa Calculator (Computer Software, retrieved from http://justus.randolph.name/kappa). The reference alpha value was 0.05. When appropriate, the Bonferroni correction was used (0.05/3 pairwise comparisons = 0.017).

Results

Study population and HRCT findings

Patients showed at least one comorbidity in 57% (44/77) of cases, and ≥2 comorbidities in 26% (20/77) of cases, respectively. Cardiovascular, oncological, and respiratory diseases were the most frequent ones, reported in 42% (32/77), 13% (10/77), and 12% (9/77) of patients, respectively. The median time period from the onset of symptoms to HRCT was 5 days (IQR, 2–9 days). 38 over 77 patients (49%) developed severe disease during the 15-day period following the HRCT examination [median (IQR) time 1 (1, 2) day].

The per-reader distribution of HRCT findings is shown in Table 1. Regardless of the reader, the most frequent features were GGO > 2 (74–83% of patients) and OP pattern (38–68% of patients), while a TLS > 6 was found in 65–69% of patients. Overall, the prevalence of HRCT features was not significantly different among readers, except for OP pattern, which was more frequently reported by R1 than R3 (52/77 versus 29/77 patients, $p < 0.001$), and by R2 than R3 (43/77 versus 29/77 patients, $p = 0.014$). Example cases are shown in Figs. 1 and 2.

Inter-reader agreement in assessing HRCT features

Table 2 shows the results of the inter-reader agreement analysis. When comparing the three radiologists at the same time, we found that they agreed to a substantial extent in assessing HCRT features (k values ranging 0.65–0.74). The highest agreement was observed in the case of GGO > 2 ($k = 0.74$) and TLS > 6 ($k = 0.69$). Exceptions were consolidation and OP pattern, for which the agreement was moderate ($k = 0.60$) and fair ($k = 0.32$), respectively.

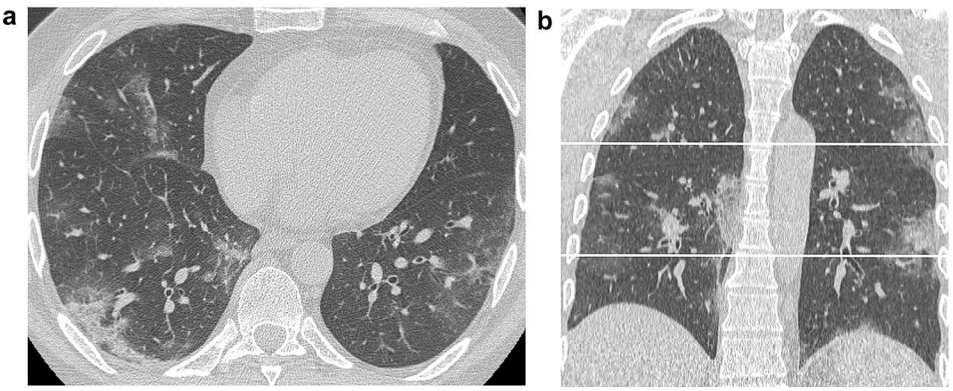
Table 1 Per-reader distribution of HRCT findings ($n=77$). The "difference in prevalence" columns report the p values expressing whether the prevalence of the detected HRCT features was significantly different among the three readers and on a pairwise basis (R1 versus R2, R1 versus R3, R2 versus R3)

HRCT feature	Prevalence of detection *n* (%, 95%CI)			Difference in prevalence			
	R1	R2	R3	Among the three readers (p)[a]	Pairwise (p)[b]		
					R	1	2
TLS > 6	50 (65, 48–86)	52 (68, 50–89)	53 (69, 52–90)	0.678	**2**	0.774	
					3	0.549	1.000
GGO > 2	57 (74, 56–96)	63 (82, 63–100)	64 (83, 64–100)	0.057	**2**	0.146	
					3	0.092	1.000
Consolidation > 2	25 (32, 21–48)	16 (21, 12–34)	22 (29, 18–43)	0.065	**2**	*0.012*	
					3	0.648	0.210
Crazy-paving pattern > 2	15 (19, 11–32)	13 (17, 9–29)	8 (10, 4–20)	0.142	**2**	0.791	
					3	0.119	0.227
OP pattern	52 (68, 50–89)	43 (56, 40–75)	29 (38, 25–54)	*<0.001*	**2**	0.035	
					3	*<0.001*	*0.014*

[a]Cochran's Q test

[b]McNemar test; HRCT, high-resolution computed tomography; 95%CI, 95% confidence interval; R1, reader 1; R2, reader 2; R3, reader 3; R, reader; TLS, total lung score; GGO, ground-glass opacity; OP, organizing pneumonia

Numbers in bold refer to R1, R2, and R3. Numbers in italic refer to p values when statistically significant



c

READERS	TFS			TLS
	GGO	Consolidation	Crazy-paving pattern	
R1	7	4	1	12
R2	6	3	1	10
R3	8	4	3	15

Fig. 1 48-year old man with confirmed COVID-19 pneumonia. At hospital admission, HRCT images on axial (**a**) and coronal (**b**) planes showed bilateral, mostly peripheral GGO and consolidations. The two horizontal white lines in (**b**) delimit the upper, middle, and lower lung zones, which were identified to apply the semi-quantitative score (see the text for details). The scheme in (**c**) resumes how each of the three readers (R1, R2, and R3) assigned the TFS for GGO, consolidation, and crazy-paving pattern, thus allowing the calculation of TLS as the sum of all the TFSs. For all the readers, TLS was >6, a feature we found to be predictive for short-term occurrence of severe disease. After one day, the patient developed respiratory failure [Italian Society of Emergency Medicine (SIMEU) phenotype III disease]

When comparing the radiologists on a pairwise basis, the agreement was moderate to substantial for most HRCT features. The only exception was the OP pattern, which was scored with moderate agreement by R1 versus R2, fair agreement by R1 versus R3, and fair agreement by R2 versus R3. Of note, the inter-reader agreement between more experienced readers (R1 and R2) was substantial for consolidation and moderate for OP pattern.

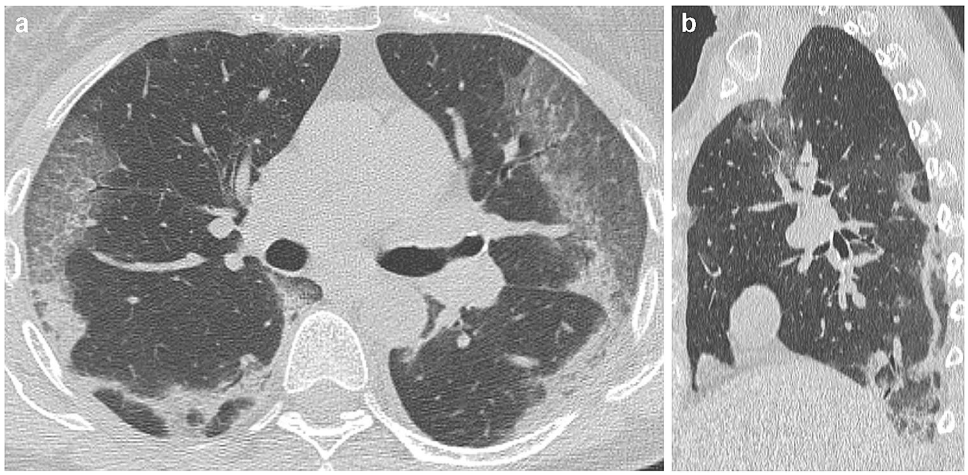


Fig. 2 61-year old woman with confirmed COVID-19 pneumonia. At hospital admission, HRCT images on axial (**a**) and sagittal (**b**) planes showed bilateral, peripheral GGO)and band-like opacities with a perilobular distribution, resembling an OP pattern. OP pattern was deemed present by all readers. After 5 days, the patient developed respiratory failure [Italian Society of Emergency Medicine (SIMEU) phenotype IV disease]

Table 2 Inter-reader agreement in assessing HRCT features. The "inter-reader agreement" columns express the magnitude of the agreement in assessing a certain HRCT feature among the three readers, and on a pairwise basis (R1 versus R2, R1 versus R3, R2 versus R3)

HRCT feature	Inter-reader agreement			
	Among the three readers *k* (95%CI)[a]	Pairwise *k* (95%CI)[b]		
		R	1	2
TLS > 6	0.69 (0.56–0.82)	**2**	0.65 (0.47–0.83)	
		3	0.68 (0.50–0.85)	0.61 (0.42–0.80)
GGO > 2	0.74 (0.62–0.86)	**2**	0.55 (0.33–0.77)	
		3	0.51 (0.28–0.73)	0.78 (0.59–0.96)
Consolidation > 2	0.60 (0.46–0.74)	**2**	0.64 (0.45–0.83)	
		3	0.42 (0.20–0.64)	0.45 (0.22–0.67)
Crazy-paving pattern > 2	0.65 (0.52–0.79)	**2**	0.64 (0.38–0.90)[c]	
		3	0.61 (0.34–0.88)[c]	0.71 (0.14–1.00)[c]
OP pattern	0.32 (0.17–0.47)	**2**	0.59 (0.42–0.77)	
		3	0.26 (0.09–0.43)	0.29 (0.10–0.49)

[a]Free-marginal multirater kappa

[b]Cohen's kappa

[c]Prevalence-adjusted bias-adjusted kappa (PABAK); HRCT, high-resolution computed tomography; 95%CI, 95% confidence interval; *R*, reader; TLS, total lung score; GGO, ground-glass opacity; OP, organizing pneumonia

Numbers in bold refer to R1, R2, and R3

Online Resource 1 shows the PA values we used to verify whether the prerequisites for using PABAK were matched (or not), as described above. PA values do not represent primary measurements of agreement, as they do not account for the effect of chance.

Prediction of unfavorable outcome

Table 3 shows the results of univariable analysis and multivariable analysis, including a model built upon each reader. On multivariable analysis, independent predictors of severe disease were TLS > 6 (for all readers), OP pattern (for R1) and consolidation > 2 (for R2).

Discussion

Chest HRCT represents a valuable imaging tool both in diagnosis and management of patients with COVID-19 [27], through suggesting a possible diagnosis of COVID-19 in a high suspicion clinical setting and indicating a progression in disease severity at follow-up (e.g., signs of disease progression such as consolidation or crazy-paving pattern, or bacterial superinfection) [28]. In this study, the agreement in assessing HRCT features of mild COVID-19 pneumonia among the three readers with different experience in thoracic imaging was fair in the case of OP pattern, moderate in the case of consolidation > 2, and substantial

Table 3 Results from the logistic regression model (outcome, development of severe disease within 15 days from HRCT)

Variable	Univariable analysis, *p*			Multivariable analysis OR (95%CI), *p*		
	R1	R2	R3	R1	R2	R3
TLS > 6	*<0.001*	*<0.001*	*<0.001*	9 (2.11–38.43), *0.003*	13.37 (3.38–52.80), *<0.001*	53.19 (6.60–428.42), *<0.001*
Consolidation > 2	*0.003*	*0.010*	*0.004*	–	4.87 (1.06–22.31), *0.042*	–
Crazy-paving pattern > 2	0.955	0.959	0.680	–	–	–
OP pattern	*<0.001*	*0.001*	0.303	6 (1.35–26.64), *0.019*	–	–

Note: OR, odds ratio; 95%CI, 95% confidence interval; R1, reader 1; R2, reader 2; R3, reader 3; TLS, total lung score; OP, organizing pneumonia

Numbers in italic refer to *p* values when statistically significant

in the case of GGO > 2, crazy-paving pattern > 2, and TLS > 6. Of note, the latter feature was found to be an independent predictor of short-term onset of severe disease at multivariable analysis, regardless of readers' experience. When considering experienced readers only, severe disease was independently predicted also by OP pattern (in the case of R1), and consolidation > 2 (in the case of R2), in accordance with higher pairwise agreement on those two features ($k = 0.59$ and 0.64, respectively). Our findings are in line with previous studies showing substantial-to-almost perfect inter-reader agreement for most CT features [14, 16], and the capability of semi-quantitative or quantitative evaluation of lung involvement [10, 12] to predict disease worsening. Overall, our results provide reliable potential markers for disease progression.

Concerning the estimation of COVID-19 pneumonia extent, Cozzi et al. recently proposed a quantitative method based on chest X-ray performed in an emergency setting, correlating with an increased risk of admission to ICU [29]. In parallel, a few Authors [11–14] evaluated lung involvement from COVID-19 pneumonia under the form of CT severity score [14], CT score [11], total lung involvement [13], and, conversely, total extent of well aerated lung parenchyma [12]. Computerized aided methods for the quantification of lung involvement in COVID-19 pneumonia were also investigated [30]. We used TLS for this purpose. This makes our results difficult to compare, though our approach shows the potential advantage of accounting for different HRCT manifestations when quantify pulmonary involvement. Of note, we found this parameter to be reliable, showing substantial inter-reader agreement for a cutoff >6, regardless of readers' experience in thoracic imaging. One can assume this is a potentially relevant result, since non-thoracic radiologists can be involved in reporting in the pandemic-related scenario. Since TLS was an independent predictor of clinical worsening, a reliable quantification of this feature, even by non-thoracic radiologists, might impact on patients' management (e.g., in terms of patients' allocation in ICU).

Concerning other HRCT features, Zhang et al. [16] found excellent inter-reader agreement in assessing consolidation ($k = 0.983$) and crazy-paving pattern ($k = 0.978$) between experienced readers. Differently from them, we observed lower (even if substantial) agreement. A potential explanation for the discrepancy might be related to the fact that our assessment included not only the presence but also the extent of those HRCT features. While this can expectedly lead to lower agreement, our approach has the potential advantage of adding a semi-quantitative evaluation of the amount of lung involvement.

The OP pattern is part of typical COVID-19 pneumonia appearance (type 1 category), according to the RSNA categorization system [7], as well as a marker of other coronavirus-related lung diseases [i.e., severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS)] [31, 32]. Consolidation has been identified as a potential marker of disease progression, reflecting cellular fibromyxoid exudates in alveoli [33, 34]. When analyzing readings from more experienced readers, both features were significantly associated to severe disease on univariable analysis (R1 and R2), with OP pattern and consolidation > 2 representing an independent predictor of the outcome for R1 and R2, respectively. However, k values for those features were disappointing, both overall and when comparing R1 or R2 versus R3. On the other hand, the agreement rose when comparing R1 and R2. Those results suggest that OP pattern and consolidation are more reliable and can represent markers for clinical evolution only when provided by experienced readers.

In the case of OP pattern, this may be due to its composite definition, which includes many HRCT findings at

a time [21]. As supported by the lower prevalence in R3 readings, the OP pattern was presumably more difficult to assess by less experienced readers. Our results concerning consolidation can be explained by the expectedly lower accuracy to assess the extent of this otherwise easier to interpret finding.

Some study limitations warrant mention. First, this is a monocentric work, suggesting that our findings should be validated by multi-institutional trials. However, our results are overall in line with previous works performed both in Western [12] and in Eastern [10, 11] scenarios of the pandemic, suggesting they are reasonably generalizable. Second, we limited our models to imaging findings, excluding patients' comorbidities and other clinical factors. On the other hand, the definition of clinically predictive models was beyond the purpose of the study. We focused on inter-reader agreement assuming that, in an emergency scenario, reliable imaging findings can represent more objective data for managing patients with known mild COVID-19 pneumonia than clinical features that can be incompletely known at the time of HRCT. Finally, we did not include the whole spectrum of HRCT features in the analysis. This choice was done to avoid asking the less experience reader to interpret subtler findings, for which the agreement would be expectedly low.

In conclusion, we observed that, regardless of radiologists' experience in chest imaging, there was moderate-to-substantial agreement for most HRCT features of COVID-19 pneumonia, in terms of semi-quantitative assessment of lung involvement. Some of the features were predictive for short-term evolution to severe disease, namely: (i) TLS > 6, showing substantial inter-reader agreement regardless of readers' experience; (ii) consolidation > 2 and OP pattern, which showed acceptable inter-reader agreement between more experienced readers only. Thus, the agreement on different HRCT features apparently depends on readers' experience.

Authors' contributions All authors contributed equally to this work.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval The Ethical Committee approved the study protocol.

Consent to participate The acquisition of the informed consent was waived, given the retrospective design.

Availability of data and material The authors confirm that the relevant data supporting the findings of this study are available within the article.

Code availability The authors confirm that the software applications used in the study are listed within the article.

References

1. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC (2020) Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA 324(8):782–793. https://doi.org/10.1001/jama.2020.12839
2. Guan WJ, Ni ZY, Hu Y et al (2020) Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 382(18):1708–1720. https://doi.org/10.1056/NEJMoa2002032
3. Corman VM, Landt O, Kaiser M et al (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 25(3):2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045
4. Inui S, Fujikawa A, Jitsu M et al (2020) Chest CT findings in cases from the cruise ship "Diamond Princess" with coronavirus disease 2019 (COVID-19). Radiol Cardiothorac Imaging 2(2):e200110. https://doi.org/10.1148/ryct.2020200110
5. Zhang JJ, Dong X, Cao YY et al (2020) Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 75(7):1730–1741. https://doi.org/10.1111/all.14238
6. Xu B, Xing Y, Peng J et al (2020) Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy. Eur Radiol 30:1–8. https://doi.org/10.1007/s00330-020-06934-2
7. Simpson S, Kay FU, Abbara S et al (2020) Radiological society of North America expert consensus statement on reporting chest CT findings related to COVID-19. Endorsed by the society of thoracic radiology, the American College of Radiology, and RSNA—secondary publication. J Thorac Imaging 35(4):219–227. https://doi.org/10.1097/RTI.0000000000000524
8. Ai T, Yang Z, Hou H et al (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 296:200642. https://doi.org/10.1148/radiol.2020200642
9. Zhou S, Wang Y, Zhu T, Xia L (2020) CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China. AJR Am J Roentgenol 214(6):1287–1294. https://doi.org/10.2214/AJR.20.22975
10. Feng Z, Yu Q, Yao S et al (2020) Early prediction of disease progression in 2019 novel coronavirus pneumonia patients outside Wuhan with CT and clinical characteristics. Nat Commun 11:4968. https://doi.org/10.1101/2020.02.19.20025296
11. Yuan M, Yin W, Tao Z, Tan W, Hu Y (2020) Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS One 15(3):e0230548. https://doi.org/10.1371/journal.pone.0230548
12. Colombi D, Bodini FC, Petrini M et al (2020) Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. Radiology 296:201433. https://doi.org/10.1148/radiol.2020201433
13. Tabatabaei SMH, Talari H, Moghaddas F, Rajebi H (2020) Computed tomographic features and short-term prognosis of coronavirus disease 2019 (COVID-19) pneumonia: a single-center study from Kashan, Iran. Radiol Cardiothorac Imaging 2:2. https://doi.org/10.1148/ryct.2020200130
14. Yang R, Li X, Liu H et al (2020) Chest CT severity score: an imaging tool for assessing severe COVID-19. Radiol Cardiothorac Imaging 2:2. https://doi.org/10.1148/ryct.2020200047
15. Li K, Fang Y, Li W et al (2020) CT image visual quantitative evaluation and clinical classification of coronavirus

disease (COVID-19). Eur Radiol 30(8):4407–4416. https://doi.org/10.1007/s00330-020-06817-6
16. Zhang R, Ouyang H, Fu L et al (2020) CT features of SARS-CoV-2 pneumonia according to clinical presentation: a retrospective analysis of 120 consecutive patients from Wuhan city. Eur Radiol 30(8):4417–4426. https://doi.org/10.1007/s00330-020-06854-1
17. Debray MP, Tarabay H, Males L et al (2020) Observer agreement and clinical significance of chest CT reporting in patients suspected of COVID-19. Eur Radiol. https://doi.org/10.1007/s00330-020-07126-8
18. Albano D, Bruno A, Bruno F et al (2020) Impact of coronavirus disease 2019 (COVID-19) emergency on Italian radiologists: a national survey. Eur Radiol 30(12):6635–6644. https://doi.org/10.1007/s00330-020-07046-7
19. Paglia S, Storti E (2020) First line Covid-19. Emergency Department Organizational Management within epidemic or pre-epidemic outbreak areas. https://www.simeu.it/w/articoli/leggiArticolo/4015/leggi/. Accessed 30 May 2020
20. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J (2008) Fleischner society: glossary of terms for thoracic imaging. Radiology 246(3):697–722. https://doi.org/10.1148/radiol.2462070712
21. Polverosi R, Maffesanti M, Dalpiaz G (2006) Organizing pneumonia: typical and atypical HRCT patterns. Radiol Med 111(2):202–212. https://doi.org/10.1007/s11547-006-0021-8
22. Feng F, Jiang Y, Yuan M et al (2014) Association of radiologic findings with mortality in patients with avian influenza H7N9 pneumonia. PLoS One 9(4):e93885. https://doi.org/10.1371/journal.pone.0093885
23. Kottner J, Audige L, Brorson S et al (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. Int J Nurs Stud 48(6):661–671. https://doi.org/10.1016/j.ijnurstu.2011.01.016
24. Cicchetti DV, Feinstein AR (1990) High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 43(6):551–558. https://doi.org/10.1016/0895-4356(90)90159-m
25. Nurjannah I, Siwi SM (2017) Guidelines for analysis on measuring interrater reliability of nursing outcome classification. Int J Res Med Sci 5(4):1169–1175. https://doi.org/10.18203/2320-6012.ijrms20171220
26. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174
27. Floridi C, Fogante M, Agostini A et al (2020) Radiological diagnosis of coronavirus disease 2019 (COVID-19): a practical guide. Acta Biomed 91(8-S):51–59. https://doi.org/10.23750/abm.v91i8-S.9973
28. Carotti M, Salaffi F, Sarzi-Puttini P et al (2020) Chest CT features of coronavirus disease 2019 (COVID-19) pneumonia: key points for radiologists. Radiol Med 125(7):636–646. https://doi.org/10.1007/s11547-020-01237-4
29. Cozzi D, Albanesi M, Cavigli E et al (2020) Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. Radiol Med 125(8):730–737. https://doi.org/10.1007/s11547-020-01232-9
30. Grassi R, Cappabianca S, Urraro F et al (2020) Chest CT computerized aided quantification of PNEUMONIA Lesions in COVID-19 infection: a comparison among three commercial software. Int J Environ Res Public Health 17(18):6914. https://doi.org/10.3390/ijerph17186914
31. Tse GM, To KF, Chan PK et al (2004) Pulmonary pathological features in coronavirus associated severe acute respiratory syndrome (SARS). J Clin Pathol 57(3):260–265. https://doi.org/10.1136/jcp.2003.013276
32. Kim I, Lee JE, Kim KH, Lee S, Lee K, Mok JH (2016) Successful treatment of suspected organizing pneumonia in a patient with Middle East respiratory syndrome coronavirus infection: a case report. J Thorac Dis 8(10):E1190–E1194. https://doi.org/10.21037/jtd.2016.09.26
33. Ye Z, Zhang Y, Wang Y, Huang Z, Song B (2020) Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. Eur Radiol 30(8):4381–4389. https://doi.org/10.1007/s00330-020-06801-0
34. Xu Z, Shi L, Wang Y et al (2020) Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med 8(4):420–422. https://doi.org/10.1016/S2213-2600(20)30076-X