CHEST

Comparison of chest CT severity scoring systems for COVID-19

Ali H. Elmokadem[1,2] · Ahmad M. Mounir[1] · Zainab A. Ramadan[1] · Mahmoud Elsedeiq[3] · Gehad A. Saleh[1]

Abstract

Purpose To compare the diagnostic performance and inter-observer agreement of five different CT chest severity scoring systems for COVID-19 to find the most precise one with the least interpretation time.

Methods and materials This retrospective study included 85 patients (54 male and 31 female) with PCR-confirmed COVID-19. They underwent CT to assess the severity of pulmonary involvement. Three readers were asked to assess the pulmonary abnormalities and score the severity using five different systems, including chest CT severity score (CT-SS), chest CT score, total severity score (TSS), modified total severity score (m-TSS), and 3-level chest CT severity score. Time consumption on reporting of each system was calculated.

Results Two hundred fifty-five observations were reported for each system. There was a statistically significant inter-observer agreement in assessing qualitative lung involvement using the m-TSS and the other four quantitative systems. The ROC curves revealed excellent and very good diagnostic accuracy for all systems when cutoff values for detection severe cases were > 22, > 17, > 12, and > 26 for CT-SS, chest CT score, TSS, and 3-level CT severity score. The AUC was very good (0.86), excellent (0.90), very good (0.89), and very good (0.86), respectively. Chest CT score showed the highest specificity (95.2%) in discrimination of severe cases. Time consumption on reporting was significantly different (< 0.001): CT-SS > 3L-CT-SS > chest CT score > TSS.

Conclusion All chest CT severity scoring systems in this study demonstrated excellent inter-observer agreement and reasonable performance to assess COVID-19 in relation to the clinical severity. CT-SS and TSS had the highest specificity and least time for interpretation.

Key Points

- *All chest CT severity scoring systems discussed in this study revealed excellent inter-observer agreement and reasonable performance to assess COVID-19 in relation to the clinical severity.*
- *Chest CT scoring system and TSS had the highest specificity.*
- *Both TSS and m-TSS consumed the least time compared to the other three scoring systems.*

Abbreviations

COVID-19	Coronavirus disease 2019
CT	Computed tomography
CT-SS	Chest CT severity score
GGO	Ground-glass opacity
m-TSS	Modified total severity score
RT-PCR	Reverse transcription-polymerase chain reaction assay
TSS	Total severity score

✉ Ali H. Elmokadem
  mokadem83@yahoo.com

1   Department of Radiology, Mansoura University, Mansoura, Egypt

2   Department of Radiology, Farwaniya Hospital, Kuwait City, Kuwait

3   Department of Anesthesia and ICU, Mansoura Faculty of Medicine, Mansoura, Egypt

Introduction

Coronavirus disease 2019 (COVID-19) has spread quickly worldwide since its initial spread in December 2019 in Wuhan, China [1]. Due to the high infection rate of the pandemic, accurate and swift diagnosis is vital to accomplish rapid and ideal management [2]. Most of the patients had mild symptoms with relatively a good prognosis, but a

minority had pulmonary edema, acute respiratory distress syndrome (ARDS), or multiple organ failure with a high mortality rate [3–5]. The mortality rate is increased in patients with ARDS and other co-morbidities such as chronic pulmonary disease, cardiovascular disease, hypertension, diabetes, and cancer [6]. The incidence of severe/critical cases was less than mild cases in multiple studies as 30.1%, 18.2%, 10.3%, and 17.6% respectively [7–10]; however, one study revealed a higher incidence of the severe disease (64.6%) [11].

The reference standard diagnostic tool of COVID-19 infection is the reverse transcription-polymerase chain reaction assay (RT-PCR) which estimates viral load from a nasopharyngeal swab or tracheal aspirate [12, 13]. Recent studies reported low sensitivity of RT-PCR in the early stage (reaching from 37 to 71%), probably due to the low viral load in test specimens or laboratory fault [14–16], while chest computed tomography (CT) has established 56–98% sensitivity in detecting COVID-19 at early presentation and can be helpful in correcting false-negative RT-PCR through the early phases of the disease [13–15]. CT chest plays an imperative role in screening, diagnosing, and evaluating the course of COVID-19 and selecting the appropriate management [17, 18]. Although chest CT has high sensitivity in COVID diagnosis, it has low specificity as it could be challenging to discriminate COVID-19 from other viral diseases on chest CT [18–20]. The chest CT abnormalities during COVID-19 are variable, and the most common changes are multifocal ground-glass opacities with or without consolidation with favorable peripheral distribution [4, 9, 19, 21, 22], including ground-glass opacities, consolidation, linear opacities, a crazy-paving pattern, and bronchial wall thickening.

Based on clinical manifestations, COVID-19 is categorized into four types: minimal, common, severe, and critical cases. Minimal disease patients have subtle symptoms. Common cases complain of fever and mild cough. Severe cases have one of these features: (1) resting blood oxygen saturation $\leq 93\%$; (2) respiratory rate ≥ 30 beats/min; or (3) oxygen concentration ≤ 300 mmHg. Critical cases have one of the following: (1) respiratory failure demanding mechanical ventilators, (2) shock, and (3) organ failure necessitating intensive care administration [4, 7].

The rapid accurate patients' categorization and radiological severity scoring are critical for appropriate management, especially in mild cases before patient deterioration; as chest X-ray has a very low sensitivity in early-stage disease, CT is the primary imaging tool [23]. Furthermore, the results of radiological examinations could be variable among radiologists, particularly in chest imaging. In order to standardize the radiological descriptions, multiple chest CT scoring systems have been developed [7–11, 17]. This study aims to compare the diagnostic accuracy and interobserver agreement of five different CT chest severity scoring systems for COVID-19, including chest CT severity score (CT-SS), chest CT score,

the total severity score (TSS), modified total severity score (m-TSS), and 3-level chest severity score in correlation with the clinical staging of disease. To the best of our knowledge, no studies have yet compared the reproducibility and interobserver agreement between these scoring systems in correlation to the clinical features and prognosis, so we aimed to detect the most reliable scoring system to save time and guide rapid, accurate management in the current pandemic.

Methods

Study population

The local institutional review board approved this retrospective study, and a waiver of the consent of medical record review was received. Ninety-two patients with PCR-confirmed COVID-19 who underwent chest CT to assess the pulmonary parenchymal severity from August 2020 to December 2020 were initially enrolled. We excluded seven patients, three patients with negative findings at chest CT, and four patients with missed clinical data. The final study cohort consisted of 85 patients classified into severe/critical and non-severe cases. Severe-cases group is presented by clinical signs of pneumonia plus one of the following: respiratory rate > 30 breaths/min; severe respiratory distress; or $SpO_2 < 90\%$ on room air-based. D-dimer values were recorded for all cases at admission, while the P/F ratio was recorded only for severe cases admitted to ICU. P/F ratio is used to assess the severity of hypoxemia and defined as the ratio of the PaO_2 (partial pressure of arterial oxygen obtained from an arterial blood gas) to the FiO_2 (fraction of inspired oxygen expressed as a decimal).

CT image acquisition

Chest CT imaging without contrast agent was done on a 16-detector CT scanner (Bright speed; GE healthcare). All patients were examined in a supine position, and images were acquired during a single inspiratory breath-hold. The scanning range was from the apex of the lung to the costophrenic angle. CT scan parameters are as follows: X-ray tube parameters, 120 KVp, 350mAs; rotation time, 0.5 s; pitch, 1.0; section thickness, 5 mm; intersection space, 5 mm; additional reconstruction with sharp convolution kernel and a slice thickness of 1.5 mm. Scans were reviewed at a window width and level of 1000 to 2000 HU and -700 to -500 HU, respectively, to assess the lung parenchyma.

Qualitative image analysis

Chest CT scans for all patients were assessed by one reviewer with 10 years of experience in thoracic imaging for

the following characteristics based on the Fleischner Society Nomenclature recommendations and similar studies [19, 24, 25]: ground-glass opacity (GGO), consolidation, nodule, crazy-paving pattern, subpleural lines, bronchial wall thickening, lymph node enlargement, and pleural effusion. The distribution of lung abnormalities was also classified as predominately peripheral or diffuse in each case.

Quantitative image analysis

To evaluate the severity of pulmonary parenchymal involvement, we attempted to quantify the extent of the abnormalities by five scoring systems. CT images were independently reviewed by three radiologists with more than 10 and 9 years of experience in thoracic imaging. Reviewers were blinded from the clinical data. Time consumption on reporting of each scoring system was calculated.

Chest CT severity score

The CT severity score (CT-SS) is an adaptation of a method used before to describe ground-glass opacity, interstitial opacity, and air trapping and was correlated with clinical and laboratory parameters in patients after SARS [10]. The 18 segments of both lungs are divided into 20 regions, in which the posterior apical segment of the left upper lobe is divided into apical and posterior segmental regions, while the anteromedial basal segment of the left lower lobe was subdivided into anterior and basal segmental regions. The lung attenuations in all 20 lung regions are subjectively evaluated on chest CT and given a score of 0, 1, or 2 if the parenchymal opacification involved 0%, less than 50%, or equal or more than 50% of each region, respectively. Thus, the CT-SS is defined as the sum of each score in the 20 lung regions, ranging from 0 to 40 points.

Chest CT score

Chest CT score is calculated per each of the 5 lobes based on the extent of parenchymal involvement [11], as follows: (0) no involvement; (1) < 5% involvement; (2) 5–25% involvement; (3) 26–50% involvement; (4) 51–75% involvement; and (5) > 75% involvement. The resulting total CT score is the sum of each individual lobar score and ranges from 0 to 25.

Total severity score

The total severity score is mainly a quantitative score assessing the inflammatory abnormalities in each of the five lobes of both lungs, including the presence of GGOs, consolidation, or mixed GGOs [7]. Depending on the percentage of the involved lobe, each lobe could be scored from 0 to 4 points: (0) = 0%, (1) = 1–25%, (2) = 26–50%, (3) = 51–75%, or (4) = 76–100%. The total score is the sum of the points from each lobe and ranges from 0 to 20.

Modified total severity score

The modified total severity score adds the character of abnormalities to the previously described total severity score (TSS) with the same score from 0 to 4 points [17]. The additional qualitative signs of lung involvement are ground-glass opacity (A), crazy-paving pattern (B), consolidations (C), and characters other than enlisted (X). The final result is the sum of the points awarded for each of the five lobes and a letter representing the predominant abnormality in both lungs.

3-level chest severity score

The extent and nature of pulmonary involvement are assessed at three levels [8]: (i) above the carina (upper level), (ii) below the carina up to the superior margin of the inferior pulmonary vein (middle level), (iii) below the inferior pulmonary vein (lower level). The extent of pulmonary involvement at each level is scored based on a 4-point scale: (0) for normal lung; (1) for < 25% lung abnormalities; (2) for 25–49% abnormalities; (3) for 50–74% abnormalities and (4) for ≥ 75% abnormalities. The nature of pulmonary involvement is evaluated from 1 to 4; (1) normal lung parenchyma; (2) at least 75% ground-glass opacities/crazy-paving pattern; (3) combination of ground-glass opacities/crazy-paving pattern and consolidation provided that each is less than 75% involvement; (4) at least 75% consolidation. The two scores (the extent and nature of pulmonary involvement) are multiplied by each other and added to the scores of all six levels (3 levels on each side). The final severity score ranges from 0 to 96.

Statistical analysis

Data were entered and analyzed by MedCalc Statistical Software version 18.9.1 (MedCalc Software bvba; http://www.medcalc.org; 2018) and IBM-SPSS version 25. Quantitative variables were expressed as means, SD, and ranges, while qualitative variables were expressed as raw numbers, proportions, and percentages. Kaplan–Meier curve was used to calculate the median survival time for ICU cases. The Fleiss' kappa test was made to estimate the inter-observer agreement between three reviewers to assess qualitative lung involvement using m-TSS. The Kappa (K) values were interpreted as follows: k values between 0.61 and 0.80 represented good agreement; k values between 0.81 and 0.90 represented very good agreement; k values between 0.91 and 1.00 represented excellent agreement. The interclass

correlation (ICC) test was done to assess the reliability in quantitative lung assessment between the three observers using the other four scoring systems. A *p* value less than 0.05 indicated a statistically significant difference. The receiver operating characteristic (ROC) curves for the pulmonary assessment using CT SS, CT severity score at three levels, chest CT score, and TSS scoring systems (including m-TSS) with a calculation of the area under the curve (AUC) were done. The m-TSS scoring system was not evaluated separately as it was considered a minor modification of the TSS. The chi-square test was done to assess the sensitivity and specificity of m-TSS in either and both quantitative and qualitative lung assessment.

Results

Patients' characteristics, clinical manifestations, and CT findings

Twenty-two (25.9%) were severe/critical cases, and 63 (74.1%) were non-severe cases. Compared with the non-severe group, the severe patients were significantly older (mean age, 58.1 years (SD, 11.1) vs. 51.8 years (SD, 15.3) $p < 0.044$). There was a statistically significantly higher respiratory rate and lower SPO_2 in severe vs. non-severe cases. The severe disease group had a significantly higher incidence of associated comorbidities like diabetes mellitus, hypertension, and ischemic heart disease. All severe cases were admitted to the ICU ($n = 22$); 13 patients were on CPAP, while nine were on mechanical ventilation. The mortality rate was 59.1% (13/22) among patients admitted to ICU. The flow chart of the study is demonstrated in Fig. 1. The median time to death (survival time) in critical cases was

96 h after ICU admission, as shown by the Kaplan–Meier curve (Supplementary Fig. 1). D-dimer values were significantly higher in severe cases versus non-severe ones (median 2.71 µg/ml [interquartile range 1.82–3.42] vs 0.56 [0.41–0.81], $z = -6.51$, $p < 0.001$) (Fig. 2). The median P/F ratio recorded for severe cases was 90 (interquartile range 74–106).

A statistically significantly higher lymph node enlargement, predominant left-sided lesions, and crazy paving pattern were found in severe versus non-severe cases, while ground-glass opacities were more frequent in non-severe cases. Characteristics of the enrolled cases are summarized in Table 1. Demonstrative non-severe and severe COVID-19 cases are shown in Figs. 3 and 4.

Inter-observer agreement

Two hundred fifty-five observations were reported for each scoring system. There was a statistically significant inter-observer agreement between the three observers in assessing qualitative lung involvement using the m-TSS (Table 2). The overall agreement was very good ($\kappa = 0.860$) for individual categories and normal findings, ground-glass opacities, and consolidations, but good for crazy paving ($\kappa = 0.786$). In addition, excellent inter-observer reliability was found among the three observers in quantitative lung assessment using the other four scoring systems CT-SS, TSS, chest CT score, and CT severity score three levels (ICC > 0.9) (Table 3 and Fig. 5).
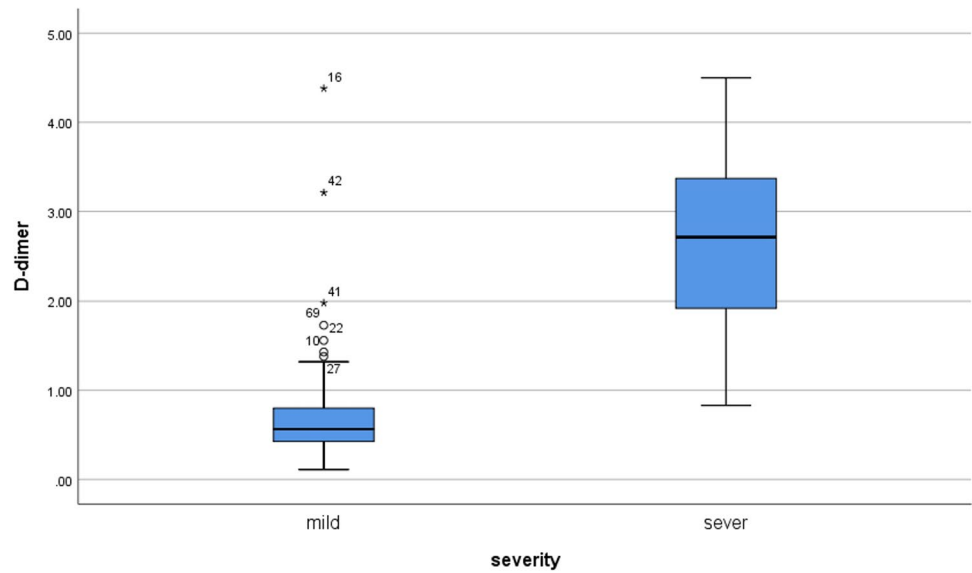
Severity scoring systems

The ROC curve was done for each scoring system separately for differentiating severe from non-severe

Fig. 1 Flow chart of the study
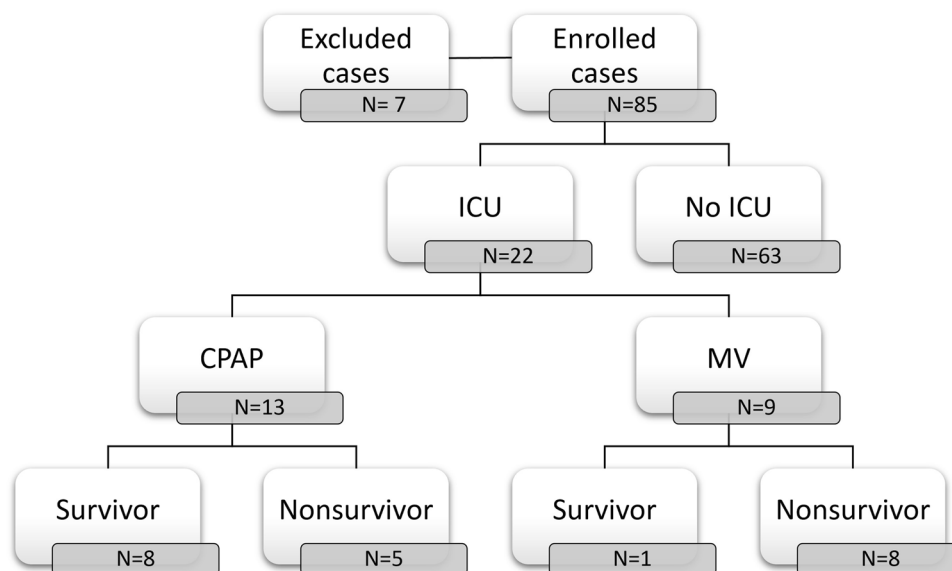
Fig. 2 Pairwise comparisons of the D-dimer values of non-sever and severe cases

Table 1 Clinical and radiological characteristics of enrolled cases

Characteristic	Total	Non-severe COVID-19	Severe COVID-19	*p* value
N	85	63	22	
Age (years)	53.4 ± 14.5	51.8 ± 15.3	58.1 ± 11.1	0.044
Sex				0.598
Male	54 (63.5%)	39 (61.9%)	15 (68.2%)	
Female	31 (36.5%)	24 (38.1%)	7 (31.8%)	
Associated comorbidities	52 (61.2%)	31 (49.2%)	21 (95.5%)	< 0.001
Diabetes	24 (28.2%)	13 (20.6%)	11 (50%)	0.008
Hypertension	29 (34.1%)	17 (27%)	12 (54.5%)	0.019
Ischemic heart disease	8 (9.4%)	2 (3.2%)	6 (27.3%)	0.003*
Chronic liver disease	9 (10.6%)	6 (9.5%)	3 (13.6%)	0.690*
SPO$_2$	89.9 ± 6.9	92.5 ± 4.7	82 ± 6.3	< 0.001$
Respiratory rate (breaths/minute)	22.1 ± 4.1	20.7 ± 2.8	26.3 ± 4.5	< 0.001$
Subpleural bands	35 (41.2%)	27 (42.9%)	8 (36.4%)	0.594
Lymph node enlargement	10 (11.8%)	4 (6.3%)	6 (27.3%)	0.017*
Dominant side				0.041
Right	64 (75.3%)	51 (81%)	13 (59.1%)	
Left	21 (24.7%)	12 (19%)	9 (40.9%)	
Distribution				0.903
Diffuse	55 (64.7%)	41 (65.1%)	14 (63.6%)	
Peripheral	30 (35.3%)	22 (34.9%)	8 (36.4%)	
Ground-glass opacities	54 (63.5%)	46 (73%)	8 (36.4%)	0.002
Consolidations	13 (15.3%)	9 (14.3%)	4 (18.2%)	0.734*
Crazy paving	10 (11.8%)	3 (4.8%)	7 (31.8%)	0.002*
Nodules	8 (9.4%)	6 (9.5%)	2 (9.1%)	1.000*

Data expression [test of significance]: *N* (%) [chi-square or *Fisher's exact test] or mean ± SD [$independent-samples *t* test]

COVID-19 cases; the 4 ROC curves revealed excellent and very good diagnostic accuracy for all scoring systems. When cutoff values for detection severe cases were > 22, > 17, > 12, and > 26 for CT-SS, chest CT score, TSS, and 3-level CT severity score, the AUC was very good (0.868), excellent (0.904), very good (0.890), and

Fig. 3 Non-contrast chest CT axial (**a**), coronal (**b**), and sagittal (**c**) images for a 40-year-old man with mild COVID-19 pneumonia. CT images show ground-glass opacities and crazy paving pattern in multiple lung segments. The CT-SS is 9, CT chest severity score is 7, TSS is 5, m-TSS is 5A, and 3-level CT severity score is 24
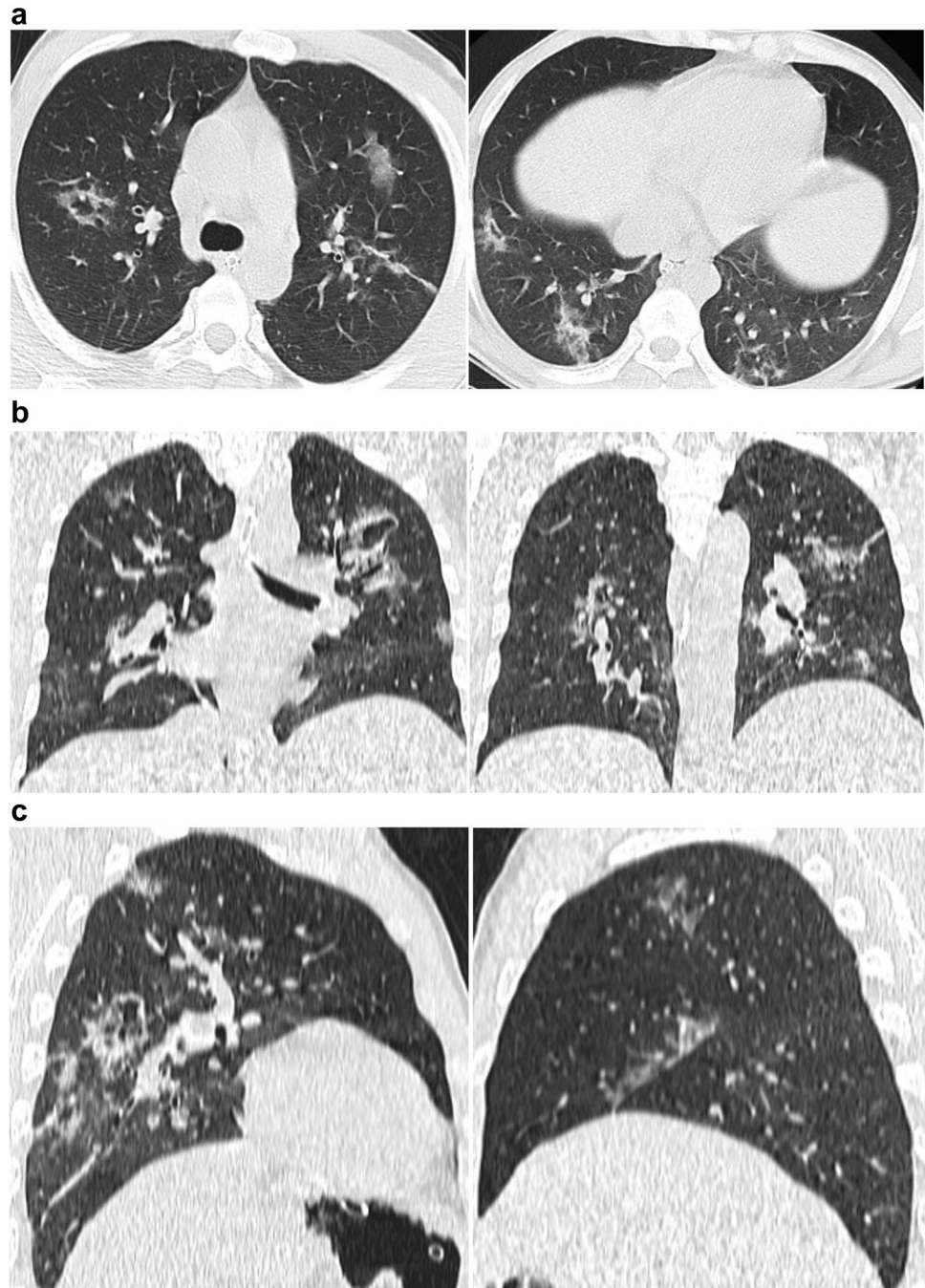

very good (0.865) respectively. Chest CT score and TSS revealed reasonable sensitivity (77.3%) with high specificity (95.2% and 90.5% respectively) in detection severe cases (Fig. 6). The comparison between these four independent ROC curves revealed no statistically significant difference between the four scoring systems.

There was a statistically significant difference in m-TSS qualitative lung scores between severe/critical patients who required ICU admission versus non-severe cases ($p < 0.001$); most of the patients who did not require ICU admission (74%) showed GGO. In comparison, most of the patients who underwent ICU admission (68.2%) showed either crazy paving (Cp) or consolidation (C) (Supplementary Table 1). Additionally, the m-TSS showed higher specificity (92%) with the cutoff value $\geq$ of 12 after the addition of the qualitative pattern, including crazy paving (Cp) and consolidation (C) changes.

Fig. 4 Non-contrast chest CT axial (**a**), coronal (**b**), and sagittal (**c**) images for a 55-year-old woman with severe COVID-19 pneumonia. CT images show ground-glass opacities and consolidation in multiple lung segments. The CT-SS is 33, CT chest severity score is 19, TSS is 16, m-TSS is 16C, and 3-level CT severity score is 72
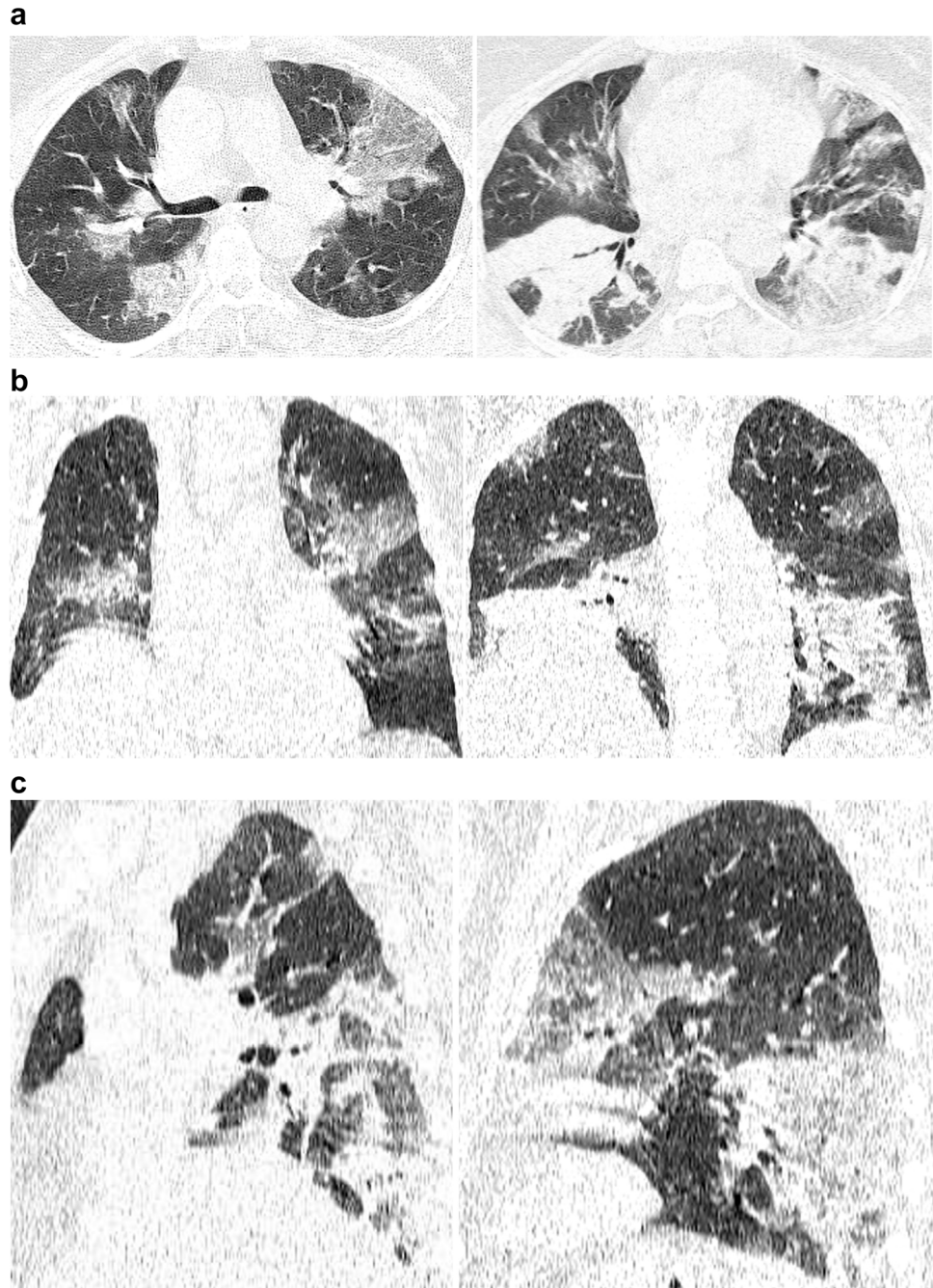
**Table 2** Inter-rater reliability test for qualitative lung involvement in m-TSS

Category	Fleiss' kappa (κ)	95% CI	SE	p value
Overall	0.860	0.776–0.945	0.043	<0.001
Normal lungs	0.825	0.702–0.948	0.063	<0.001
Ground glass opacities	0.868	0.745–0.991	0.063	<0.001
Crazy paving	0.786	0.663–0.908	0.063	<0.001
Consolidations	0.907	0.785–1.030	0.063	<0.001

SE asymptotic standard error. Test of significance: Fleiss' kappa (κ)

Table 3 Inter-rater reliability test for quantitative scoring systems

Scoring system	ICC	95% CI	p value
CT severity score (CT-SS)	0.991	0.987–0.994	<0.001
Total severity score (TSS)	0.994	0.992–0.996	<0.001
Chest CT score	0.993	0.989–0.995	<0.001
CT severity score three levels	0.987	0.982–0.991	<0.001

ICC intraclass correlation coefficient, *CI* confidence interval. Test of significance: scale reliability analysis
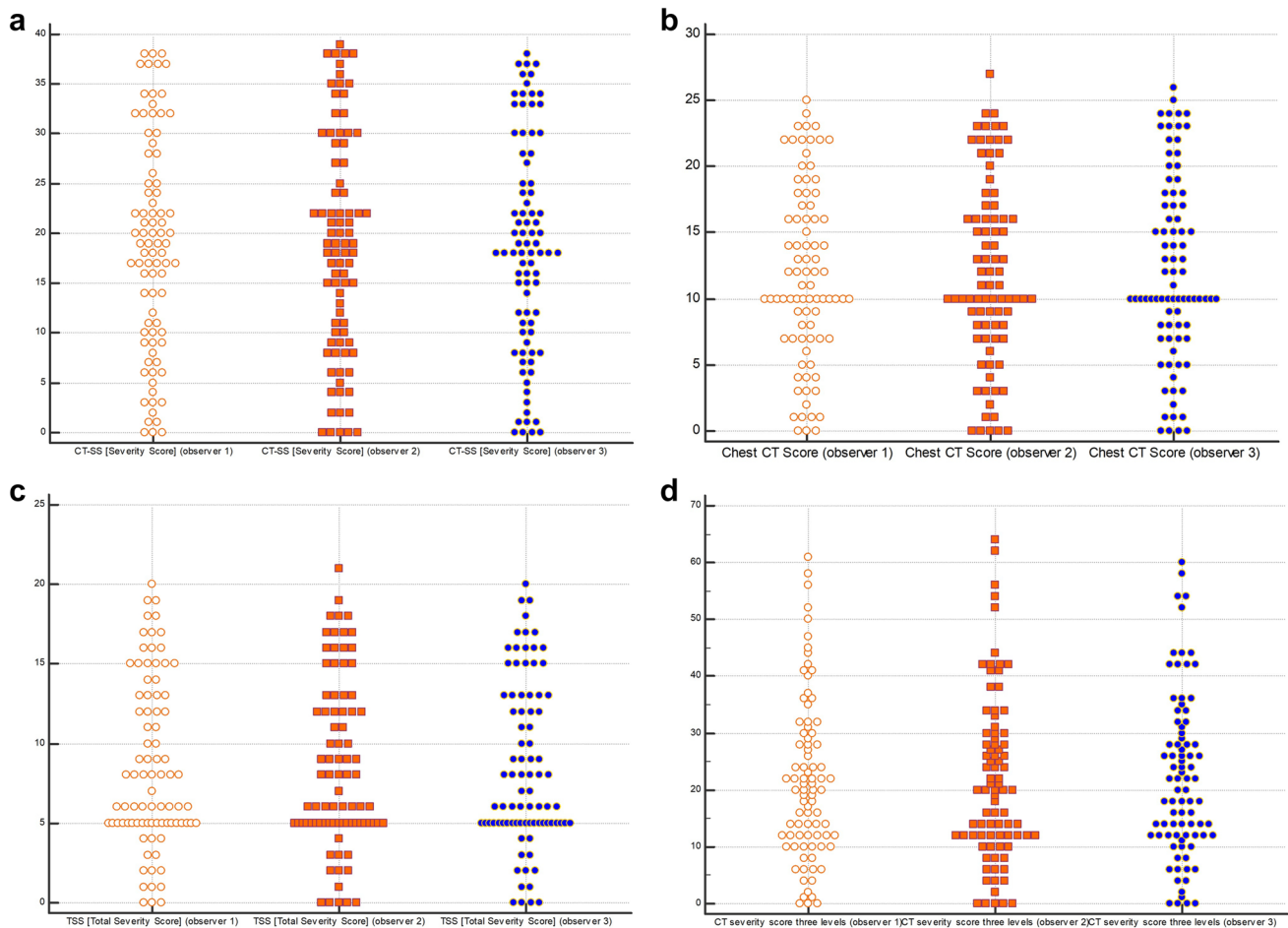
Fig. 5 Multiple dot graphs for inter-observer agreement for CT-SS (**a**), CT chest severity score (**b**), TSS (**c**), and 3-level CT severity score (**d**)

Time consumption

Time consumption on reporting of each scoring system was calculated under the same reading environment and using similar diagnostic monitors. Kruskal–Wallis *H*-test revealed a statistically significant difference (<0.001) in scoring time: CT-SS > 3-level CT severity score > Chest CT score > TSS (Table 4). Furthermore, pairwise comparisons showed a statistically significant difference between all pairs except CT SS vs. CT SS three levels (Fig. 7).

Discussion

As COVID-19 has rapidly spread worldwide, many scoring systems have been published for pulmonary assessment. In this retrospective study, we conducted a comparative study of five CT scoring systems correlated with clinical manifestation and prognosis. There was a statistically significant inter-observer agreement between three independent observers for the overall evaluation of the pulmonary

abnormalities in COVID-19 patients using the m-TSS scoring system. Similarly, there was excellent reliability in lung assessment using the other four scoring systems CT-SS, TSS, chest CT score, and CT severity score three levels (ICC > 0.9). A similar design was adopted in a recent case–control study that compared the performance and interobserver agreement four diagnostic scoring systems: COVID-19 Reporting and Data System (CO-RADS), the COVID-19 imaging reporting and data system (COVID-RADS), the RSNA expert consensus statement, and the British Society of Thoracic Imaging (BSTI) [26]. Unlike our study, there was no correlation with the clinical implications of these systems and the diagnosis of COVID-19; also the authors of the current study investigated involvement of the lung with different severity scores, while the other studies investigated the diagnostic performance of different diagnostic scoring systems.

Our results were concordant with prior studies that reported inter-observer reliability of the severity scoring systems. The inter-reader agreement for CT-SS was excellent in two different studies (ICC median = 0.925,
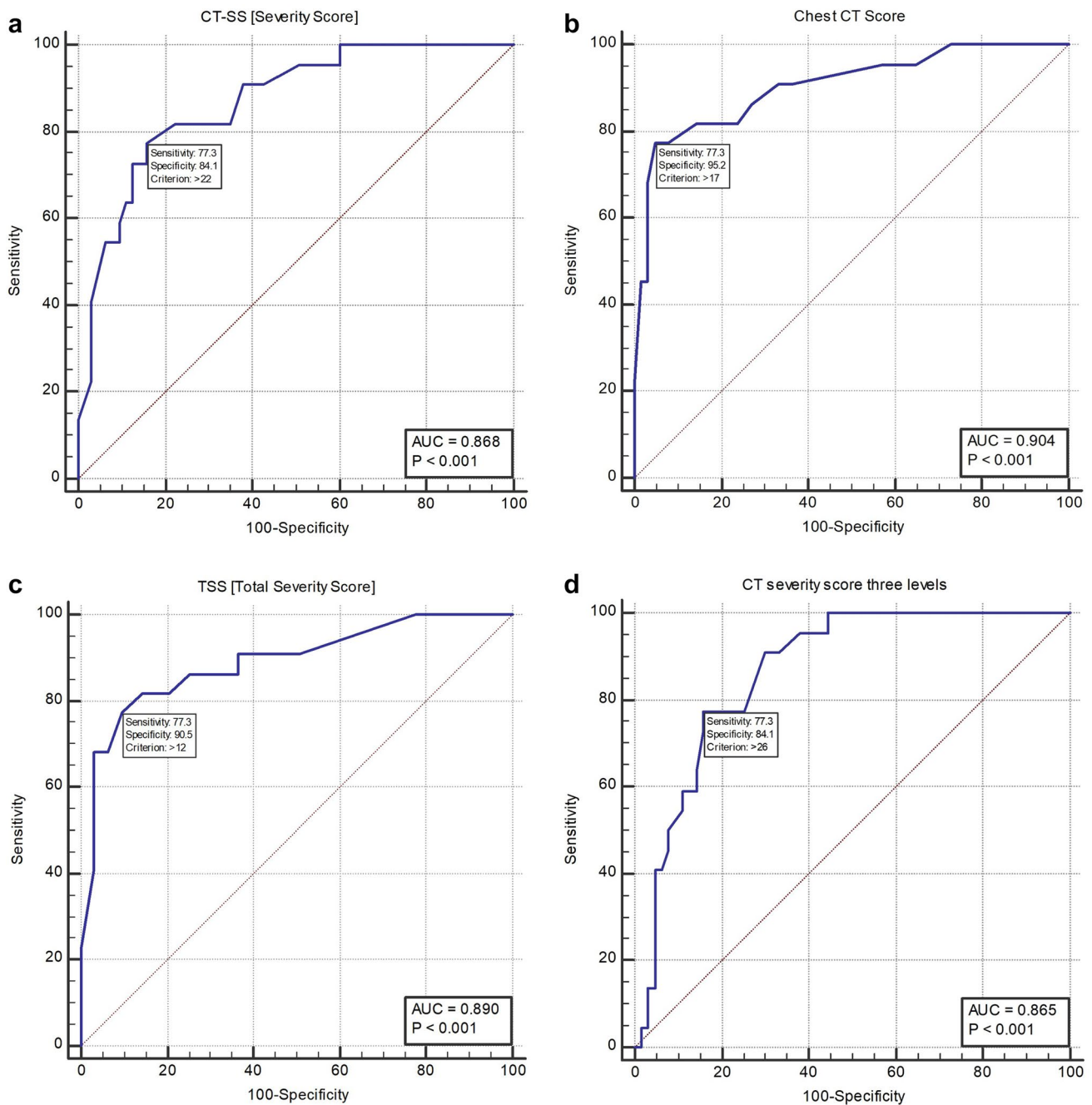
Fig. 6 ROC curves for diagnostic performance of CT-SS (**a**), CT chest severity score (**b**), TSS (**c**), and 3-level CT severity score (**d**) in detection of severe cases

ICC mean = 0.936 and K = 0.85, $p = 0.001$) [10, 27]. Li et al reported excellent inter-observer consistency of the CT visual quantitative analysis with ICC 0.976 (95% CI 0.962–0.985) between 2 observers using only the TSS [7]. Similarly, the inter-observer agreement of 2 readers was excellent in a study performed to assess the 3-level severity scoring system (intra-class correlation coefficient 0.908, 95% CI 0.882–0.931;

$p < 0.001$) [8]. Chest CT scoring system was correlated with clinical and laboratory status of the COVID-19 patients but the inter-observer agreement was not performed [11].

The m-TSS scale is an update of the TSS where additional qualitative features of pulmonary abnormalities were added [17]; however, the system was not evaluated by inter-observer reliability or correlated with clinical severity. As

Table 4 Comparisons of interpretation time (minutes) of each of the 4 scoring systems

Statistic	Chest CT score	TSS	CT SS	CT SS 3L	p value
Median	11	9	14	13	<0.001
IQR	9–12	8–10	12–16	11–14	
Range	5–16	4–12	6–19	5–17	
Pairwise comparisons	A	B	C	C	

p value: Kruskal–Wallis *H* test. Pairwise comparisons: similar letters = insignificant difference, different letters = significant difference

regards the m-TSS, the overall agreement was very good (κ = 0.860). The inter-observer agreement was also very good for individual categories but good for crazy paving (κ = 0.786). In this study, the CT imaging features were reliable with the previous literature reports [22, 28–30] as most of the patients had GGO and mixed GGO with consolidations of multifocal peripheral or diffuse distribution. Our study revealed a statistically significantly higher crazy paving pattern and a statistically significantly lower ground-glass opacities in severe vs. non-severe cases; the same prevalence has been reported in many previous studies [14, 22, 31, 32]. However, one study revealed no statistical incidence difference in GGO detection between the two groups [9]. The frequency of GGOs detected in non-severe cases primarily denotes the correlation between the imaging of the acute-phase diffuse alveolar damage and airspace edema [33], while the frequency of crazy-paving pattern in severe cases possibly states a mixture of alveolar edema, bacterial

superinfection, and interstitial inflammatory changes [34, 35].

Prior studies were performed to assess the diagnostic accuracy of each system, but no studies compared the diagnostic accuracy among scoring systems. All scoring systems in this study demonstrated excellent and very good diagnostic accuracy when cutoff values for detection severe cases were > 22, > 17, > 12, and > 26 for CT-SS, chest CT score, TSS, and 3-level CT severity score. Our results showed a slightly less sensitivity and higher specificity of the chest CT scoring system (77.3% and 95.2%, respectively) compared to the previous study, which revealed sensitivity and specificity of 80.0% and 82.8% for discriminating critical and mild cases [9]. Additionally, Francone et al reported significantly higher chest CT scores in critical than in mild-stage patients and among late-phase than early-phase patients ($p < 0.0001$). Chest CT score was significantly correlated with CRP ($p < 0.0001$, $r = 0.6204$) and D-dimer ($p < 0.0001$, $r = 0.6625$) levels. Similar to our results, a CT score of ≥ 18 was associated with increased mortality risk [11]. Another study reported a significantly higher median TSS of the severe-type group as compared to the common type ($p < 0.001$) and a cutoff value of 7.5 to have 82.6% sensitivity and 100% specificity [7] compared to 77.3% sensitivity and 90.5% specificity when using a cutoff value of 12 in the current study. A ROC analysis for 3-level CT severity score revealed 38 as a cutoff value for predicting the development of critical symptoms with a sensitivity of 93.33%, a specificity of 59.26%, and an area under the curve (AUC) of 0.843 (95% CI 0.778–0.895; $p < 0.0001$) [8].

Fig. 7 Pairwise comparisons of the interpretation time of the severity scoring systems
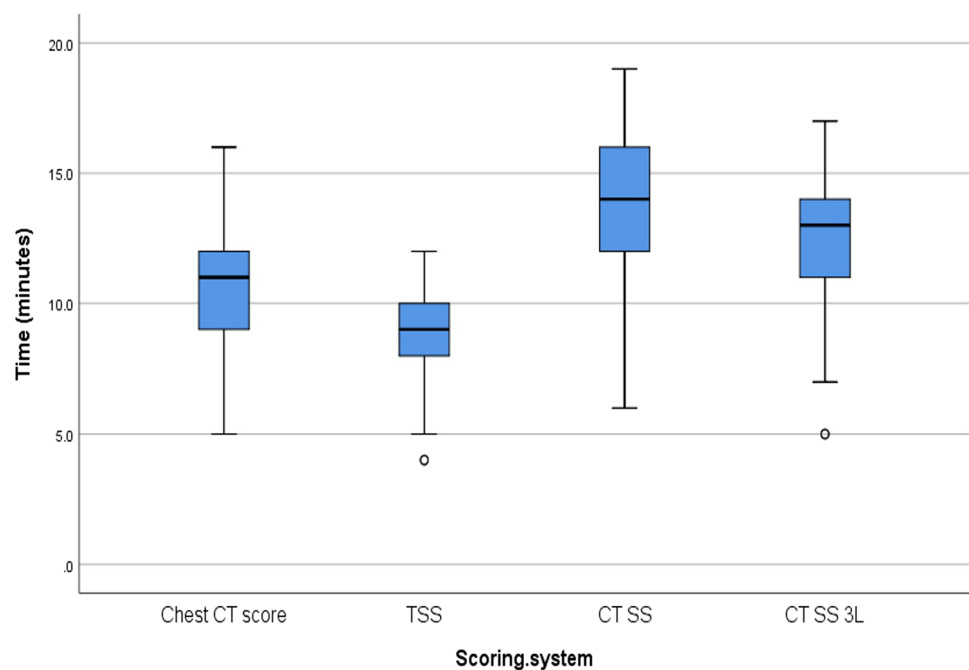
Kaplan–Meier curve in this study shows that the median time to death (survival time) for ICU cases was 96 h after ICU admission. The critical/severe cases were less than (25.9%) mild cases and showed a relatively high mortality rate (59.1%). In concordance with our results, multiple recent studies reported a worse prognosis and higher mortality rate among patients with severe/critical COVID-19 disease than mild/typical disease [7, 11, 23].

Reducing the interpretation time needed for severity scoring is a great consideration for a busy radiology department, especially after adding the burden of the COVID-19 pandemic. The pulmonary assessment using both TSS and m-TSS consumed the least time (average 10 min) compared to the other three scoring systems. CT-SS consumed the longest time for interpretation as it requires segmental assessment, which means smaller regions and more intervals to consider during evaluation. 3-level severity score also consumed a longer time for interpretation as it requires assessment of the extent and nature of parenchymal lesions separately and multiplication of the results to get the final score.

This study has few limitations. First, its retrospective design relatively limits the identification of the prognostic factors. Secondly, we revealed excellent reproducibility compared to other studies; this may be due to the single-center design of the study, the use of a single CT scanner, and strict application of laboratory-confirmed COVID-19 cases; these variables are assumed to have favorably influenced image interpretation. Thirdly, the two groups were not balanced in so far as the group with severe/critical disease was relatively small. Further studies with more patients, particularly severe patients, are needed. Fourthly, there was no exact information about when the symptoms began and when CT was acquired. Lastly, none of our patients underwent a lung biopsy to imitate the histopathological changes. Future studies comparing the performance of artificial intelligence, machine learning or deep-learning-based tools, and CT-assisted pulmonary software against radiologist-based severity scoring systems in terms of clinical operability, time consumption, and accuracy are recommended.

Conclusion

Severity scoring has a great implication for the precise diagnosis, management, and follow-up of COVID-19 cases. All chest CT severity scoring systems in this study evaluated the severity of COVID-19 with an excellent inter-observer agreement and reasonable performance. Chest CT scoring system and TSS had the highest specificity and least time for interpretation. We recommend using severity scoring systems as a part of the standard report of chest CT for COVID-19 patients.

Declarations

References

1. Zhu N, Zhang D, Wang W et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382(8):727–733
2. Mahase E (2020) Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 368:m1036
3. Grasselli G, Pesenti A, Cecconi M (2020) Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA 323(16):1545–1546
4. Chen N, Zhou M, Dong X et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395(10223):507–513
5. Xu Y-H, Dong J-H, An W-M et al (2020) Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. J Infect 80(4):394–400
6. Guzik TJ, Mohiddin SA, Dimarco A et al (2020) COVID-19 and the cardiovascular system: implications for risk assessment, diagnosis, and treatment options. Cardiovasc Res 116(10):1666–1687
7. 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
8. Salaffi F, Carotti M, Tardella M et al (2020) The role of a chest computed tomography severity score in coronavirus disease 2019 pneumonia. Medicine (Baltimore) 99(42):e22433
9. Li K, Wu J, Wu F et al (2020) The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. Invest Radiol 55(6):327–331
10. Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q et al (2020) Chest CT severity score: an imaging tool for assessing severe COVID-19. Radiol Cardiothorac Imaging 2(2):e200047

11. Francone M, Iafrate F, Masci GM, Coco S, Cilia F, Manganaro L et al (2020) Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol 30(12):6808–6817

12. Wong HYF, Lam HYS, Fong AH et al (2020) Frequency and distribution of chest radiographic findings in patients positive for COVID-19. Radiology 296(2):E72–E78

13. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML et al (2020) Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. Radiology 296(2):E46-54

14. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P et al (2020) Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 296(2):E115–E117

15. Kanne JP, Little BP, Chung JH, Elicker BM, Ketai LH (2020) Essentials for radiologists on COVID-19: an update-radiology scientific expert panel. Radiology 296(2):E113–E114

16. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J (2020) Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing. Radiology 296(2):E41–E45

17. Wasilewski PG, Mruk B, Mazur S, Półtorak-Szymczak G, Sklinda K, Walecki J (2020) COVID-19 severity scoring systems in radiological imaging–a review. Pol J Radiol 85:e361

18. Ai T, Yang Z, Hou H et al (2020) Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 296(2):E32–E40

19. Bernheim A, Mei X, Huang M et al (2020) Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. Radiology 295(3):200463

20. Elmokadem AH, Batouty NM, Bayoumi D, Gadelhak BN, Abdel-Wahab RM, Zaky M et al (2021) Mimickers of novel coronavirus disease 2019 (COVID-19) on chest CT: spectrum of CT and clinical features. Insights Imaging [Internet] 12(1):12. Available from: https://doi.org/10.1186/s13244-020-00956-6

21. Elmokadem AH, Bayoumi D, Abo-Hedibah SA, El-Morsy A (2021) Diagnostic performance of chest CT in differentiating COVID-19 from other causes of ground-glass opacities. Egypt J Radiol Nucl Med 52(1):1–10

22. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X et al (2020) CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology 295(1):202–207

23. Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel coronavirus outbreak of global health concern. Lancet 395(10223):470–473

24. 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

25. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J (2008) Fleischner Society: glossary of terms for thoracic imaging. Radiology 246(3):697–722

26. Inui S, Kurokawa R, Nakai Y, Watanabe Y, Kurokawa M, Sakurai K et al (2020) Comparison of chest CT grading systems in coronavirus disease 2019 (COVID-19) pneumonia. Radiol Cardiothorac Imaging. 2(6):e200492

27. Abo-Hedibah SA, Tharwat N, Elmokadem AH (2021) Is chest X-ray severity scoring for COVID-19 pneumonia reliable? Pol J Radiol 86:e432–e439

28. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A (2020) Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. AJR Am J Roentgenol 215(1):87–93

29. Wong HYF, Lam HYS, Fong AH-T, Leung ST, Chin TW-Y, Lo CSY et al (2020) Frequency and distribution of chest radiographic findings in patients positive for COVID-19. Radiology 296(2):E72-8

30. Pan F, Ye T, Sun P et al (2020) Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19). Radiology 295(3):715–721

31. Xiong Y, Sun D, Liu Y et al (2020) Clinical and high-resolution CT features of the COVID-19 infection: comparison of the initial and follow-up changes. Invest Radiol 55(6):332–339

32. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J et al (2020) Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 20(4):425–434

33. Li X, Geng M, Peng Y, Meng L, Lu S (2020) Molecular immune pathogenesis and diagnosis of COVID-19. J Pharm Anal 10(2):102–108

34. Koo HJ, Lim S, Choe J, Choi S-H, Sung H, Do K-H (2018) Radiographic and CT features of viral pneumonia. Radiographics 38(3):719–739

35. Tian S, Xiong Y, Liu H, Niu L, Guo J, Liao M et al (2020) Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. Mod Pathol 33(6):1007–1014