Research Article

Comparing Visual and Software-Based Quantitative Assessment Scores of Lung Parenchymal Involvement Quantification in COVID-19 Patients

Marco Nicolò[1,2], Altin Adraman[3], Camilla Risoli[4], Anna Menta[2], Francesco Renda[5], Michele Tadiello[2], Sara Palmieri[2], Marco Lechiara[2], Davide Colombi[4], Luigi Grazioli[2], Matteo Pio Natale[6], Matteo Scardino[7], Andrea Demeco[8], Ruben Foresti[8], Chiara Martini[9]

1. ASST Spedali Civili di Brescia, Brescia, Italy; 2. Department of Diagnostic Imaging, Spedali Civili di Brescia, Brescia, Italy; 3. Unit of Neuroradiology, University-Hospital of Padova, Padova, Italy; 4. Department of Radiological Functions, Radiology Unit, Guglielmo da Saliceto Hospital, Piacenza, Italy; 5. ASST Rhodense, Italy; 6. Department of Respiratory Disease, University of Foggia, Italy; 7. Department of Diagnostic Imaging and Interventional Radiology, A.O.U. Citta della Salute e della Scienza di Torino, Torino, Italy; 8. Department of Medicine and Surgery, University of Parma, Italy; 9. Department of Health Professions, University of Parma, Italy

Background. Computed tomography (CT) plays a paramount role in the characterization and follow-up of Covid-19. Several scoring systems have been implemented to properly assess the lung parenchyma involved in patients suffering from Sars-Cov-2 infection, such as visual quantitative assessment score (VQAS) and software-based quantitative assessment score (SBQAS).

Purpose. This study aims to compare VQAS and SBQAS with two different software.

Material and methods. This was a retrospective study; 90 patients were enrolled with the following criteria: patients' age more than 18 years old, positive test for Covid-19, and unenhanced chest CT scans obtained between March and June 2021.

The VQAS was independently assessed, and the SBQAS was performed with two different Artificial Intelligence-driven softwares (Icolung and CT-COPD). The Intraclass Correlation Coefficient (ICC) statistical index and Bland-Altman test were employed.

Results. The agreement score between radiologists (R1 and R2) for the VQAS of the lung parenchyma involved in the CT images was good (ICC = 0.871). The agreement score between the two software applications for the SBQAS was moderate (ICC = 0.584). The accordance between Icolung and the

median of the visual evaluations (Median R1-R2) is good (ICC = 0.885). The correspondence between CT-COPD and the median of the VQAS (Median R1-R2) is moderate (ICC = 0.622).

Conclusion. This study showed moderate and good agreement regarding the VQAS and the SBQAS, enhancing this approach as a valuable tool to manage Covid-19 patients.

Introduction

Radiological imaging played a crucial part during the coronavirus disease 2019 (Covid-19) pandemic. Computed tomography (CT) played a paramount role in the characterization and follow-up of the illness, and its importance is broadly accepted [1][2].

Typical manifestations of Covid-19 pneumonia on chest CT images have been reported in various studies [3][4], such as ground-glass opacity (GGO), which is a non-specific term defined by the Fleischner Society as the presence on high-resolution computed tomography (HRCT) of a hazy increase in lung density, not associated with obscuration of the underlying vessels or bronchial walls; when vessels are obscured, the proper term used is "consolidation" [5].

Various studies investigated the possibility of drafting a tailored low-dose chest CT protocol for infected patients, such as Homayounieh F et al [6], which discussed this matter through a survey issued by the International Atomic Energy Agency (IAEA) from May to July 2020. The questionnaire collected data regarding scan parameters, dose-related information, whether a dedicated CT protocol for Covid-19 patients was in place, how many CT scanners were available in the facility, and which type of CT protocol was most used for these patients. The authors analyzed CT acquisition protocols across all vendors. It resulted that a limit of CTDIvol (Volume CT Dose Index) less than 3 mGy (Gray) is acceptable when the evaluation is limited to the lung parenchyma. Besides, they encouraged the use of iterative reconstruction to achieve a lower dose for infected patients. A systematic review made by Suliman, I.I. et al [7] collected different low-dose chest CT protocols for Covid-19 through varied scientific databases. The authors gathered the scanning parameters from the main papers comparing the standard protocol (STD) versus the ultra-low-dose one (ULD). It was enhanced as follows: lower kV, pitch higher than 1, use of iterative reconstruction (IR), tube current modulation, and fixed mAs were implemented to achieve the ULD protocol.

A number of software programs had been developed during the pandemic to help radiologists in the diagnosis of Covid-19, especially when lung CT scans were the most requested exams.

These software programs showed their utility to face the increased workload and to accelerate the process of diagnosis. Connected to these software [8], several score systems had been implemented to properly assess the lung parenchyma involved in patients suffering from Sars-Cov-2 infection. They have been mainly divided into two methods: visual quantitative assessment score (VQAS) and software-based quantitative assessment score (SBQAS). The first one relies on the amount of lung abnormalities visually recognized by experienced radiologists, while the second one is built upon software based on artificial intelligence (AI) to automatically or semi-automatically detect lesions and give a report about the quantification of lung parenchyma involved.

Therefore, this study aims to compare the visual quantitative and software-based assessment between two different software packages regarding the quantification of lung parenchyma affected by SARS-CoV-2 infection, in order to investigate the differences and strong points within them and to establish their reliability.

Material and Methods

Study Population

Approval for this study was granted by the local ethics committee (approval number NP5928). The institutional review board waived the requirement to obtain written informed consent for this retrospective case series, since all analyses were performed on de-identified data; therefore, there was no potential risk to patients.

Ninety patients were included with the following criteria: patients' age more than 18 years old, real-time polymerase chain reaction (RT-PCR) test positive for COVID-19, and an unenhanced chest CT scan obtained between March and June 2021. All patients that did not meet these criteria were not included in the study.

Age, gender, weight, height, BMI (Body Mass Index), and clinical indication for chest CT were recorded at the time of the examination.

doi.org/10.32388/QSC6ZD

CT Protocol

The entire population of this study underwent a chest CT scan without the injection of a contrast agent on a 64-detector scanner (Philips Brilliance 64; Amsterdam, The Netherlands).

The scanning range was from the apex to the base of the lungs, with the images obtained at full inspiration in the supine position. The chest CT parameters were as follows: kV range 100-140 kV, 80-350 effective mAs, using both z-axis and angular tube current modulation, fixed mAs 30-80 for a few patients (n=11) [7], 0.4 s rotation time, and pitch 0.8 to 1.2 (Table 1). All data were reconstructed using a sharp reconstruction kernel for parenchyma evaluation and the constructor's iterative reconstruction iDose[4] with a strength of 4 to 7. Window center and window width were set at -600; -1600. As it can be seen, there were no dedicated COVID-19 parameters for chest CT scans, resulting in the use of different strategies to achieve proper dose and image quality, such as fixed mAs and a higher level of IR [7].

Technical Parameters	
kV	100-120-140
effective mAs	80-350
fixed mAs	30-80
rotation time	0.4
pitch	0.8-1.2
individual detector size	0.625 mm
detector configuration	64x0.625 mm
thickness	2 mm
increment	2 mm

Table 1. shows the technical parameters used to acquire the chest CT scans.

Radiation doses were expressed in Computed Tomography Dose Index (CTDI) and Dose-Length-Product (DLP). Mean and median CTDI were respectively 8.23±4.20 and 7.13, while mean and median

DLP were respectively 383.3±208.88 and 342.

Visual quantitative assessment score

The VQAS for each patient in this study was made independently by two radiologists (S.P and M.L) with more than 10 years of experience.

CT images were independently reviewed and analyzed according to the Fleischner Society Glossary of Terms for Thoracic Imaging [9]. The reviewers were also blinded from the clinical data, and they categorized CT findings as highlighted by Sverzellati et al [10].

The VQAS was formulated according to some previous studies [11][12][13]. In particular, the two readers gave a percentage as a result of the lung parenchyma involved by Covid-19 following the criteria of the total severity score proposed by Li K. et al [14].

Software-based quantitative assessment score

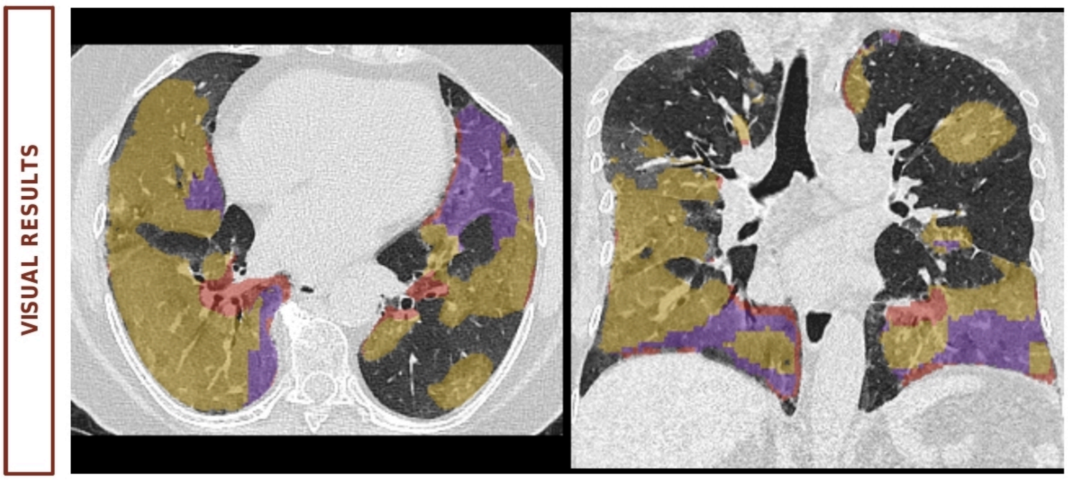
The software-based assessment score (SBQAS) was performed with two different AI-based software.

The first one, "Icolung" (Icometrix, Leuven, Belgium), is a cloud-based software that automatically contours the lungs. Moreover, it returns a report with the percentage of the lung parenchyma involved [15]. This software is based on deep learning (DL) models that sequentially carry out fully automated lung segmentation and identify abnormalities, such as ground-glass opacity (GGO), crazy-paving pattern (CPP), and consolidation. The report shows the abnormalities visualized in 2D axial and coronal views and a table with the total lung involvement percentages, divided for each lobe, and the corresponding severity scores (0-5 score per each lobe) based on Pan, F. et al. [16].

The second software used in this study is called Philips IntelliSpace Portal clinical application CT-COPD (Philips, Eindhoven, The Netherlands) computer tool. It is a semi-automatic software for lung segmentation; it was mainly used to measure the extent of chronic obstructive pulmonary *disease* (COPD). It enables the setting of a threshold of unit Hounsfield (HU) to quantify the lung parenchyma according to the needs of the operator. In this study, the HU threshold chosen to establish the lung parenchyma affected by SARS-CoV-2 infection was set at -750, as proposed in other studies [8][17][18]. The SBQAS for this tool was performed by a trained radiographer (A.M), and the result was obtained by considering the percentage of total lung parenchyma amount minus the extent of the percentage of aerated residual lung volume.

Examples of reports obtained from Icolung and CT-COPD are respectively illustrated in Figure 1 and Figure 2.

icolung icometrix



INVOLVEMENT

Affected %*	Total	GGO	CO	CPP
Left upper lobe	20 %	14 %	1 %	4 %
Left lower lobe	56 %	38 %	5 %	12 %
Right upper lobe	18 %	17 %	<1 %	<1 %
Right middle lobe	56 %	43 %	3 %	9 %
Right lower lobe	72 %	50 %	8 %	13 %
Total lungs	**40 %**	**29 %**	**3 %**	**7 %**

* ● GGO - Ground Glass Opacity, ● CO - Consolidation, ● CPP - Crazy Paving Pattern

Figure 1. It shows an example of a report from Icolung (Icometrix, Leuven, Belgium)

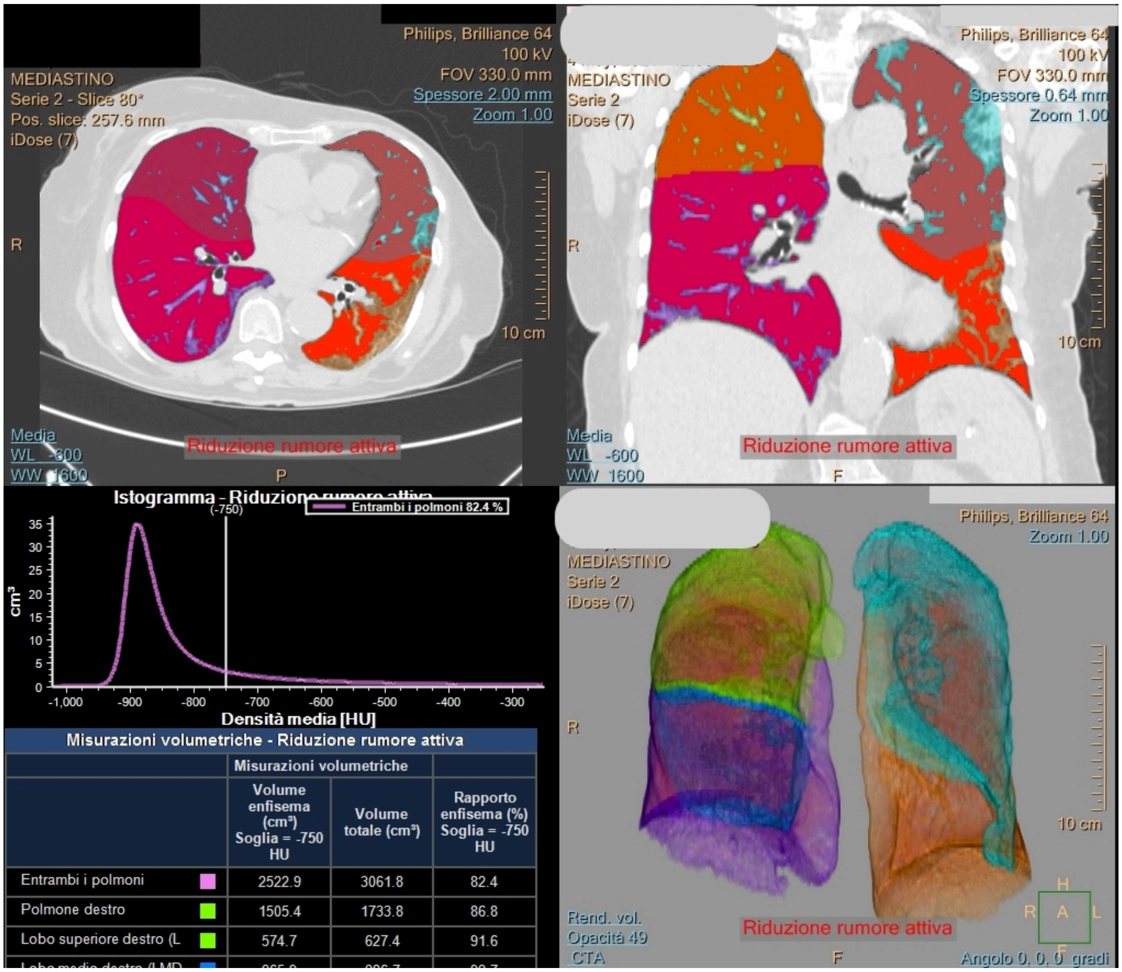


Figure 2. It displays an example of a report from CT-COPD (Philips, Eindhoven, The Netherlands)

Statistical Analysis

The statistical analysis was conducted using IBM SPSS (Statistical Package for the Social Sciences) version 29.0.1.0 {m/171/} and Prism GraphPad version 9.5.1 to ensure comprehensive data analysis and accurate interpretation. Categorical variables were presented as counts and percentages, while continuous variables were expressed as medians.

To assess agreement among the two radiologists regarding the VQAS for lung parenchyma involvement on the CT images, as well as between the software quantification, the Intraclass Correlation Coefficient (ICC) statistical index was employed. Interpreting the ICC values, the following criteria were utilized: values less than 0.50 were indicative of poor reliability, values ranging from 0.50 to 0.75 indicated moderate reliability, values ranging from 0.75 to 0.90 indicated good reliability, and

 doi.org/10.32388/QSC6ZD

values greater than 0.90 indicated excellent reliability. This interpretation framework helped assess the level of agreement and reliability achieved in both the radiologists' visual quantitative assessment score for disease extension and the software-based assessment score of lung parenchyma involved.

By utilizing the ICC, it was possible to quantitatively evaluate the degree of agreement and reliability among the raters or assessments, providing valuable insights into the consistency and concordance of their evaluations.

Results

Patient Characteristics

79 patients were considered, with a mean age of 69 ± 12 years, ranging from a minimum of 37 to a maximum of 95. The interquartile range (IQR, 25° and 75°) was, respectively, 59 and 78 years.

Inter-Reader Agreement of Visual and Software-Based CT Assessment

The agreement between radiologists (R1 and R2) for the visual quantitative assessment score of the lung parenchyma involved in the CT images was good (ICC = 0.871). The agreement between the two software packages (Icolung and CT-COPD) for the SBQAS is moderate (ICC = 0.584).

The descriptive statistics and the boxplot of the two software packages and the radiologists are summarized, respectively, in Table 2 and Figure 3.

	25TH PERCENTILE	**MEDIAN**	**75th Percentile**
ICOLUNG	8	18	29,5
CT-COPD	30	42,7	62,5
R1	10	30	50
R2	5	15	35

Table 2. Summary of descriptive statistics of the two software (Icolung and CT-COPD) and the two radiologists (R1 and R2).

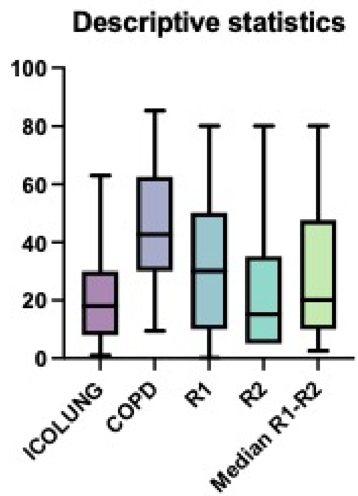


Figure 3. Boxplot scheme of the two software (Icolung and CT-COPD), the two radiologists (R1 and R2), and the median of R1-R2

The agreement between Icolung and the median of the visual evaluations (Median R1-R2) is good (ICC = 0.885). The agreement between CT-COPD and the median of the visual evaluations (Median R1-R2) is moderate (ICC = 0.622).

Interestingly, the second software, CT-COPD, has an overestimation of the results, as indicated by the higher median, first and third percentiles. Also, the first radiologist (R1) presents higher values (median, first and third percentiles) as regards to the second one (R2) (Figure 4).

doi.org/10.32388/QSC6ZD

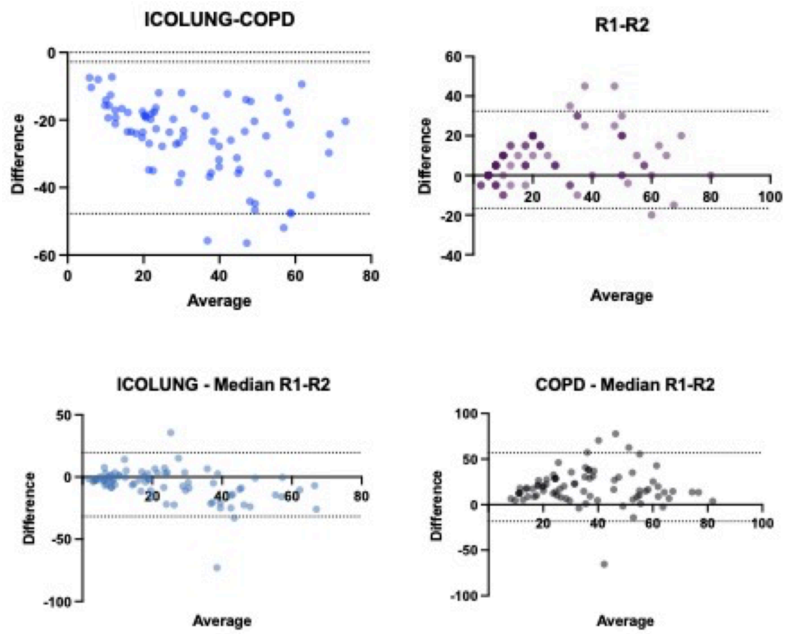


Figure 4. Bland-Altiman graphics show the trend of the values assessed. The first graph represents the assessment between the two software, Icolung and CT-COPD; results lie in a range between -47.73 to -2.78 with an SD of ±1.96. The second represents the comparison between the visual descriptions of radiologists (R1 and R2); results lie in a range between -16.59 to 32.44 with an SD of ±1.96. The third graph shows the first software (Icolung) vs. the median of the visual estimations; results lie in a range between -31.75 to 19.59 with an SD of ±1.96. The fourth graph shows the second software (CT-COPD) vs. the median of the visual estimations; results lie in a range between -18.41 to 56.75 with an SD of ±1.96.

Finally, 5 patients were excluded from the statistical analysis due to massive breathing artifacts because they could lead the software to a miscalculation of the lung parenchyma involved, and the software's output results were labeled as outliers for 6 patients, so they were excluded as well.

Discussion

Managing Covid-19 patients by assessing the extent of the lung parenchyma involved was cardinal during the Covid-19 pandemic. Hence, AI resulted in a valid and helpful tool to assist physicians in this process.

This study has shown good agreement (ICC = 0.871) between the two-blinded radiologists (R1 and R2) for the visual quantitative assessment score of the lung parenchyma involved in the CT images. This indicates a univocal method of lung parenchyma abnormalities detection. Besides, the agreement between the two software systems (Icolung and CT-COPD) for the SBQAS is moderate (ICC = 0.584). This result could be explained by analyzing the nature of these two different software systems. Icolung is an automatic DL software trained during the pandemic, while CT-COPD is a software designed to quantify chronic obstructive pulmonary disease and adapted to evaluate the extension of the lung parenchyma affected by SARS-CoV-2. In addition to this, it was found that the agreement between Icolung and the median of the visual evaluations (Median R1-R2) is good (ICC = 0.885). The agreement between CT-COPD and the median of the visual evaluations (Median R1-R2) is moderate (ICC = 0.622). This aspect outlines the validation of the AI-based software. Overall, it is worth mentioning that CT-COPD presents an overestimation of the results, as indicated by the higher median, first, and third percentiles. This may rely on the possibility of editing the lung parenchyma contouring proposed by the software.

The topic of this article has been investigated by several authors, each with different peculiarities. Granata V. et al [17] used the clinical application CT-COPD (Philips, Eindhoven, The Netherlands) to evaluate the critical lung involvement in patients vaccinated or unvaccinated affected by different variants of SARS-CoV-2, finding this tool suitable for pathological abnormalities, mainly regarding the assessment of consolidation. Besides, they obtained good statistically significant correlations among volumes extracted by the automatic tool for each lung lobe and the overall radiological severity score. Another study conducted by Durhan G. et al [18] retrospectively assessed Covid-19 patients who underwent chest CT. The authors compared the VQAS with the normal lung parenchyma percentage made by a DL software and suggested that this parameter could give valuable and objective information about pneumonia due to the infiltrative nature of lung involvement.

DL software implemented in radiology to evaluate patients affected by SARS-CoV-2 has been used since the Covid-19 outbreak. Sab L. et al [19] compared six different AI paradigms, and the authors demonstrated that AI can automatically extract tissue features and characterize the disease, distinguishing between non-Covid-19 pneumonia and Covid-19 pneumonia. Clinical examples of these models can be found in other studies, such as Suri J.S. et al [20] and Gujot J. et al [21], in which the earlier-cited software offered a valid tool to detect and classify affected patients. Nevertheless, Jungmann F. et al [22] stated their concern regarding the actual AI solutions, such as Icolung, as tools to assess positive predictive value (PPV). This article emphasized a low and variable specificity and low positive predictive value of AI solutions investigated in detecting Covid-19 pneumonia in chest CT. Finally, the authors suggested carefulness in using such tools to avoid false-positive patients. This current study overcame what was suggested by the earlier article by enrolling patients with a positive RT-PCR for Covid-19 disease.

One of the main advantages of automatic or semi-automatic involved parenchyma quantification is to help stratify patients when it comes to admission into the hospital, as it could result in lowering the cost. For example, Caruso D. et al [13] suggested using quantitative chest CT integrated with clinical parameters to help accurately triage Covid-19 patients. Additionally, Esposito G. et al [15] proposed the Icolung tool (Icometrix, Leuven, Belgium) as a practical tool to flag high-risk patients and lower healthcare costs. The authors analyzed the transmission of Sars-Cov-2 infection, expressed as cost per avoided infection, and the in-hospital length of stay of Covid-19 patients, expressed as cost per avoided hospital days, creating a framework that may allow physicians to make decisions on hospital policy and resource allocation.

Therefore, as far as our knowledge extends, this is the first study that compares an automatic AI-driven lung segmentation tool and a semi-automatic one with the visual quantitative assessment score made by radiologists.

After all, this article presents a few limitations. Firstly, the retrospective nature of the study. Secondly, the number of patients enrolled (n=90) and the engagement of just one operator to perform the SBQAS with the semi-automatic software. Lastly, the HU threshold set for Icolung to detect lung abnormalities is different from the one used by CT-COPD due to the nature of the deep learning process.

In conclusion, this study showed moderate and good agreement between the VQAS and the SBQAS produced by the two software programs and the two radiologists. Therefore, this type of AI software could be used as a reliable tool to assist the diagnostic process.

Acknowledgments

The authors thank all the colleagues who actively acquired the chest images and assisted the radiological process during the pandemic for their remarkable work.

References

1. ^Mikeal M.; Boike S.; Benedict T. et al, The Role of Computed Tomography in the Management of Hospitalized Patients With COVID-19, Cureus, 2023;28:15-e36821
2. ^Marrie-Pierre R.; Parkar AP.; Prosch H. et al, COVID-19 patients and the radiology department - advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI), Eur Radiol. 2020;30:4903-4909
3. ^Cozzi D.; Cavigli E.; Moroni C. et al. Ground-glass opacity (GGO): a review of the differential diagnosis in the era of COVID-19, Jpn J Radiol 2021;39:721-732
4. ^Pontone G.; Scafuri S.; Mancini ME. et al. Role of computed tomography in COVID-19. J Cardiovasc Comput Tomogr. 2021;15:27-36
5. ^Hansell DM.; Bankier AA.; MacMahon H. et al Fleischner Society: glossary of terms for thoracic imaging. Radiology 2008;246:697-722
6. ^Homayounieh F.; Holmberg O.; Umairi RA. et al, Variations in CT Utilization, Protocols, and Radiation Doses in COVID-19 Pneumonia: Results from 28 Countries in the IAEA Study. Radiology 2021;29:E141-E151
7. a, b, cSuliman, I.I.; Khouqeer, G.A.; Ahmed, N.A. et al. Low-Dose Chest CT Protocols for Imaging COVID-19 Pneumonia: Technique Parameters and Radiation Dose. Life (Basel). 2023;12:13-992
8. a, bRisoli, C.; Nicolò, M.; Colombi, D. et al, Different Lung Parenchyma Quantification Using Dissimilar Segmentation Software: A Multi-Center Study for COVID-19 Patients. Diagnostics 2022, 12, 1501
9. ^Hansell, D.M.; Bankier, A.A.; MacMahon, H. et al. Fleischner Society: Glossary of Terms for Thoracic Imaging. Radiology 2008;246:697–722
10. ^Sverzellati, N.; Milanese, G.; Milone, F. et al, Integrated Radiologic Algorithm for COVID-19 Pandemic. J. Thorac. Imaging 2020;35:228–233
11. ^A. H. Elmokadem; A. M. Mounir; Z. A. Ramadan, et al. Comparison of chest CT severity scoring systems for COVID-19, Eur Radiol. 2022;32:3501–3512
12. ^Colombi, D.; Petrini, M.; Maffi, G. et al, Comparison of admission chest computed tomography and lung ultrasound performance for diagnosis of COVID-19 pneumonia in populations with different disease prevalence, Eur J. Radiol, 2020, 133,109344
13. a, bCaruso, D.; Zerunian M.; Polici M. et al, Diagnostic performance of CT lung severity score and quantitative chest CT for stratification of COVID-19 patients, Radiol Med 2022;127:309–317

14. ^Li K.; Fang Y.; Li W. et al (2020) CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID- 19). Eur Radiol 2020;30:4407–4416
15. a, bEsposito, G.; Ernst, B.; Henket, M. et al. AI-Based Chest CT Analysis for Rapid COVID-19 Diagnosis and Prognosis: A Practical Tool to Flag High-Risk Patients and Lower Healthcare Costs. Diagnostics (Basel). 2022;1:12-1608
16. ^Pan, F.; Ye, T.; Sun, P. et al. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology 2020;295:715–721
17. a, bGranata, V.; Fusco, R.; Villanacci, A. et al. Imaging Severity COVID-19 Assessment in Vaccinated and Unvaccinated Patients: Comparison of the Different Variants in a High Volume Italian Reference Center. J. Pers.Med. 2022;12:955
18. a, bDurhan G, Ardalı Düzgün S, Başaran Demirkazık F, et al. Visual and software-based quantitative chest CT assessment of COVID-19: correlation with clinical findings. Diagn. Interv Radiol 2020;26:557–564
19. ^Saba L, Agarwal M, Patrick A, et al. Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs, Int J of Comput Assist Radiol Surg 2021;16:423–434
20. ^Suri, J.S.; Agarwal, S.; Chabert, G.L. et al. COVLIAS 2.0-cXAI: 7 Cloud-Based Explainable Deep Learning System for COVID-19 Lesion Localization in Computed Tomography Scans. Diagnostics (Basel). 2022;12:1482
21. ^Guiot, J.; Vaidyanathan, A.; Deprez, L. et al. Development and Validation of an Automated Radiomic CT Signature for Detecting COVID-19. Diagnostics (Basel) 2020;11:41
22. ^Jungmann, F., Müller, L., Hahn, F. et al. Commercial AI solutions in detecting COVID-19 pneumonia in chest CT: not yet ready for clinical implementation? Eur Radiol. 2022;32:3152–3160

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.