



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A new standard scoring for interstitial pneumonia based on quantitative analysis of ultrasonographic data: A study on COVID-19 patients

Fiorella Anna Lombardi^{a,*}, Roberto Franchini^a, Rocco Morello^a, Ernesto Casciaro^a, Stefania Ianniello^b, Maurizio Serra^c, Francesco Satriano^c, Francesco Mojoli^{d,e}, Silvia Mongodi^e, Daniela Pignatelli^a, Marco Di Paola^a, Sergio Casciaro^a

^a National Research Council – Institute of Clinical Physiology, Lecce, Italy

^b Diagnostic Imaging Unit, National Institute for Infectious Diseases "L. Spallanzani" IRCCS, Rome, Italy

^c Pneumology Unit 2, Vito Fazzi Hospital, ASL Lecce, Lecce, Italy

^d Department of clinical-surgical, diagnostic and pediatric sciences, Unit of anesthesia and intensive care, University of Pavia, Pavia, Italy

^e Anesthesia and Intensive Care, Fondazione IRCCS Policlinico S. Matteo, Pavia, Italy

ARTICLE INFO

Keywords:

Pneumonia
SARS-CoV-2
Lung ultrasound
Pneumonia score
Lung staging
COVID Index

ABSTRACT

Objective: To assess the effectiveness of 3 novel lung ultrasound (LUS)-based parameters: *Pneumonia Score* and *Lung Staging* for pneumonia staging and *COVID Index*, indicating the probability of SARS-CoV-2 infection.

Methods: Adult patients admitted to the emergency department with symptoms potentially related to pneumonia, healthy volunteers and clinical cases from online accessible databases were evaluated. The patients underwent a clinical-epidemiological questionnaire and a LUS acquisition, following a 14-zone protocol. For each zone, a *Pneumonia score* from 0 to 4 was assigned by the algorithm and by an expert operator (kept blind with respect to the algorithm results) on the basis of the identified imaging signs and the patient *Lung Staging* was derived as the highest observed score. The output of the operator was considered as the ground truth. The algorithm calculated also the *COVID Index* by combining the automatically identified LUS markers with the questionnaire answers and compared with the nasopharyngeal swab results.

Results: Overall, 556 patients were analysed. A high agreement between the algorithm assignments and the expert operator evaluations was observed, both for *Pneumonia Score* and *Lung Staging*, with the latter having sensitivity and specificity over 92% both in the discrimination between healthy/sick patients and between sick patients with mild/severe pneumonia. Regarding the *COVID Index*, an area under the curve of 0.826 was observed for the classification of patients with/without SARS-CoV-2.

Conclusion: The proposed methodology allowed the identification and staging of patients suffering from pneumonia with high accuracy. Moreover, it provided the probability of being infected by SARS-CoV-2.

1. Introduction

Pneumonia usually involves the outer non-mediastinal pleural surface and progresses through stages. Lung ultrasound (LUS) is a robust imaging technique able to detect pulmonary changes associated with pneumonia and their evolution depending on the degree and extent of consolidation [1]. Indeed, it has been widely used as a non-invasive bedside technique for identifying, diagnosing and following-up pneumonia and other respiratory diseases [2–7]. Although its capability to provide relevant diagnostic information on pulmonary tissue has been investigated since the 1990s [8,9], the standardization of a procedure

for the quantitative scoring of disease severity is still debated [10–12].

In 2020, the usefulness of LUS emerged during the pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [13,14]. The improvement in sensitivity in the early identification of the associated coronavirus disease (COVID-19) when integrating LUS with reverse transcription polymerase chain reaction (RT-PCR) has been recently demonstrated [15]. Moreover, both for the generic pneumonia management and in the case of COVID-19, LUS has been presented as a first-line diagnostic imaging, alternative to chest computed tomography (CT), particularly useful in children, pregnant women, critical patients and patients in areas with high rates of community transmission [13,16,

* Corresponding author.

E-mail address: fiorella@ifc.cnr.it (F.A. Lombardi).

<https://doi.org/10.1016/j.rmed.2021.106644>

Received 3 June 2021; Received in revised form 14 September 2021; Accepted 5 October 2021

Available online 7 October 2021

0954-6111/© 2021 Elsevier Ltd. All rights reserved.

17]. Additionally, CT waiting lists are particularly long, making the employment of CT for the identification of COVID-19 patients logistically challenging [18], besides being intrinsically inappropriate because of the involved radiation, which also hinders its application for the short-term follow-up.

On the other hand, LUS has not yet been widely adopted among the routine clinical procedures to face COVID-19 pandemic, probably due to the lack of automatic and objective approaches, which in fact results in the need for expert sonographers for scan execution and data interpretation.

One of the main limitations to an intensive LUS use in the triage phase is actually the need of experienced operators able to recognize the characteristic pneumonia patterns, and consequently return its classification and staging [19]. Therefore, it would be of utmost importance to provide the medical community with an advanced technology for the automatic and operator-independent identification and staging of pneumonia using quantitative LUS.

A novel technology based on the automatic analysis of ultrasonographic data has been implemented aiming to recognize the characteristic pneumonia patterns, without requiring experienced sonographers. Actually, the system guides the operator during the lung echographic scan, which can involve up to 14 anatomical zones, and automatically generates a report indicating, for each zone, the level of pneumonia severity on a scale from 0 to 4 (where 0 indicates the absence of the disease and 4 represents the maximum severity).

The aim of this study was to introduce this new methodology and to preliminarily investigate its usefulness and accuracy in pneumonia evaluation and staging, by addressing two main primary objectives: (i) assessing the accuracy of a novel LUS-measured parameter (*Pneumonia Score*) in the scoring of pneumonia in each analysed point with respect to the corresponding scoring performed by an expert operator; (ii) assessing the effectiveness of a derived parameter (*Lung Staging*) in the total staging of pneumonia, again with respect to an expert operator. As secondary objective, the performance of a further novel index (*COVID Index*), related to the probability that the observed pneumonia was due to SARS-CoV-2 infection, was evaluated.

2. Materials and methods

2.1. Study design and setting

This study was performed in collaboration with the Emergency and Admission Department (DEA) of the “Vito Fazzi” Hospital in Lecce (Italy), the Spallanzani Hospital in Rome (Italy) and the San Matteo Hospital in Pavia (Italy). The recruited patients were both healthy volunteers and patients with lung infections of various aetiology. The patients underwent a LUS scan using the *SensUS Touch – ECOVIDUS* version device (Amolab Srl, Lecce, Italy), equipped with the *Lung 19* software module and provided in an “open” configuration, specific for research purposes.

The enrolment period lasted from April 15th to May 31st 2020. To reach the required sample size, as detailed later in text, additional clinical cases were obtained from clinical databases freely accessible online, such as Butterfly, Grepmed and ThePocusAtlas (available at butterflynetwork.com, grepmed.com and thepocusatlas.com, respectively).

The resulting dataset was randomly split into two equal-sized sub-datasets, which were separately used: the first to implement the algorithm with the clinical parameters described below (cf. “Outcome” paragraph) and the second as an independent cohort to validate their performance.

The protocol was approved by the institutional review board. All patients provided an informed consent and all data were immediately anonymized. The study was conducted in accordance with the principles of the Declaration of Helsinki for clinical research involving human subjects.

2.2. Selection of participants

The eligibility criteria for the inclusion in the study were: both women and men; all ethnicities; age >18 years; healthy volunteers or subjects who referred to the point of care for triage because of one or more of the following conditions: (i) Travel/residence history in the areas with a high rate of COVID-19 transmission; (ii) Exposure to patients with fever or respiratory symptoms who were in the areas with a high rate of COVID-19 transmission; (iii) Epidemiological associations with COVID-19 infection; (iv) Clinical manifestations of COVID-19 infection such as fever, cough, hypoxemia or other respiratory symptoms, radiographic features of pneumonia (such as ground glass opacity or patchy consolidation in the lungs), normal or decreased number of white blood cells, decreased lymphocyte count in the early stage of disease, etc.

2.3. Measurements

The recruited patients underwent a clinical and epidemiological questionnaire prior to the ultrasound acquisition, outlined as an adaptation of the World Health Organization seroepidemiological investigation protocol for COVID-19 infection [20], in order to collect information about concomitant symptoms (*i.e.*: fever; sore throat; respiratory symptoms including cough, difficulty in breathing and wheezing; loss or alteration of smell and taste; diarrhea) and likelihood of exposure to SARS-CoV-2 (*i.e.*: travel or residence in a country with local transmission in the 14 days prior to the onset of symptoms; close contact with probable or confirmed cases in the 14 days prior to the onset of symptoms).

The device used to perform the LUS scan consisted of a battery based ultra-portable ultrasonographic unit and a convex probe operating at the nominal frequency of 3.5 MHz. The *Lung 19* software module integrated in the device included a novel proprietary algorithm for the automatic recognition of the characteristic patterns of pneumonia, whose working principle is detailed in the next paragraph. The LUS acquisition was guided by the software, which had most of the presets pre-configured with locked values (*e.g.*, tissue harmonics off). The only unlocked parameters were depth, focus and gain: depth (ranging from 90 to 210 mm) was set according to the patient’s BMI; focus was set as close as possible to the pleural line level; gain value was selected by the operator to obtain the better visualization of the pleura and possible pneumonia markers. In general, the locked preset values combined with the indicated setting of depth, focus and gain resulted in a quick and effective acquisition procedure, also in presence of potentially challenging situations (*e.g.*, obese patients). A water-soluble and hypoallergenic coupling gel, *e.g.* Aquasonic® 100 Ultrasound Gel (Parker Laboratories, Fairfield, NJ, USA), was used for the probe-skin coupling. The probe was suitably preserved from contamination through the use of a latex probe cover. Moreover, the employed portable ultrasound device was easily sanitized before and after each use with disinfectant wipes.

The LUS acquisition protocol was fully guided by a software interface and was based on the 14-zones method [21], namely 7 zones per side (left and right). However, because one of the main scopes of the adopted methodology is to perform a quick diagnosis, a minimum of four zones was considered sufficient in the triage phase, although the acquisition of more zones was allowed and left to the clinician judgement. The location and the minimum number of zones were selected according to recent studies already available in the literature [22]. Each acquired zone was scanned with the probe in longitudinal and transversal direction, with 12 frames acquired per scan (thus, a total of 24 frames per zone were acquired). In detail, for each patient, the following zones could be acquired: (1) Posterior portion of the lower quadrant of the right lung; (2) Posterior portion of the middle quadrant of the right lung; (3) Posterior portion of the upper quadrant of the right lung; (4,5,6) Same as (1,2,3) for the left lung; (7) Sub-axillary portion of the lower quadrant of the right lung; (8) Sub-axillary portion of the upper quadrant of the right

lung; (9,10) Same as (7,8) for the left lung; (11) Anterior portion of the lower quadrant of the right lung; (12) Anterior portion of the upper quadrant of the right lung; (13,14) Same as (11,12) for the left lung. A snapshot of the software interface after the acquisition is reported in Fig. 1. The same zones and the same images were evaluated by both the algorithm and the expert operator, in order to make the assigned scores fully comparable.

2.4. Outcomes

The algorithm automatically analysed each frame through specific morphological filters and thresholds based on the geometrical distributions of the pixels in each image in order to automatically recognize the presence of the following signs: focal, multi-focal, confluent B-lines or "lung comets"; small or consistent consolidations (in particular: small multifocal, intralobular or interlobular with possible dynamic aerial bronchogram consolidations); A-lines; pleural effusions.

The working principle of the proposed method is shown in Fig. 2. For each acquired zone on the patient chest, the algorithm assigns a *Pneumonia Score* value through the procedure described below.

For each considered anatomical zone, each image is segmented through the following steps (applied to either longitudinal and transversal acquisitions and illustrated in Fig. 3a for the case of a transversal image):

1. Preliminary image validation, based on grey level and geometrical feature analysis, in order to verify the image suitability for the subsequent processing steps and to discard possible images of insufficient quality.
2. Image pre-processing, consisting in grey level adjustment and image cutting and resizing (Fig. 3b).
3. Search for 'raw' target structures (*i.e.*, anatomical landmarks, echographic markers and signs), based only on pixel cluster positions and their grey level intensity values, where the term 'raw' indicates that this first attempt of marker segmentation may necessitate further refinements because the identified pixel clusters could contain some imperfections due to background noise (Fig. 3c).
4. Identified 'raw' target structures undergo dedicated processing steps based on a series of image filtering procedures and morphological operations, including in particular median filtering, pixel erosion/

dilatation, and hole filling, in order to finally detect the sought target structures (Fig. 3e-f).

5. Storing of the coordinates of the target structures, in order to obtain their dimensions and relative positions (Fig. 3g).

Once all the images belonging to the considered zone underwent the segmentation process through the above listed steps, in the parameter extraction phase the algorithm compares all the dimensions and relative positions resulting from step 5 and calculates the characteristic features of the detected structures (*e.g.*, number of B-lines, width of the B-lines and of the white lung, dimensions of the consolidations, etc.). The *Pneumonia Score* associated to the examined anatomical zone is computed from these parameters based on the criteria reported later in this paragraph. The maximum of the *Pneumonia Score* values obtained on a given patient is then labelled as *Lung Staging*. The outcomes of the parameter extraction phase, combined with the information deriving from the questionnaire, are finally used to estimate the *COVID Index* as detailed later in the text.

For each analysed zone, the *Pneumonia Score*, ranging from 0 to 4, was assigned by the algorithm on the basis of the identified pneumonia signs (see also Fig. 4). This score expressed the following disease grades: "0" in case of no disease, with presence of A-lines and absence of vertical artefacts (Score 0 in Fig. 4); "1" in case of early stage of the disease, with less than 3 focal B-lines (Score 1 in Fig. 4); "2" in case of intermediate stage of disease, with diffused and multifocal B-lines (Score 2 in Fig. 4); "3" in case of advanced stage of the disease, with diffused and multifocal B-lines and initial subpleural consolidations and white lung pattern (Score 3 in Fig. 4); "4" in case of very severe stage of the disease, with diffuse and multifocal B-lines and lung consolidations (Score 4 in Fig. 4). Then, the *Lung Staging* was obtained as the highest *Pneumonia Score* identified during the examination, and consequently it ranged from 0 to 4 as well. This diagnostic parameter represented the overall staging of the disease as a global assessment of the lung health status. Of note, the adopted software was thought and designed to give the operator the possibility to correct the *Pneumonia Score* values assigned by the algorithm. However, this function was inhibited in the device used for the present study, since the purpose was to validate the automatic algorithm with respect to the scores assigned by the expert operator, employing a "double-blind" approach. For this reason, changes to the automatically assigned scores were never made in any step of this study.

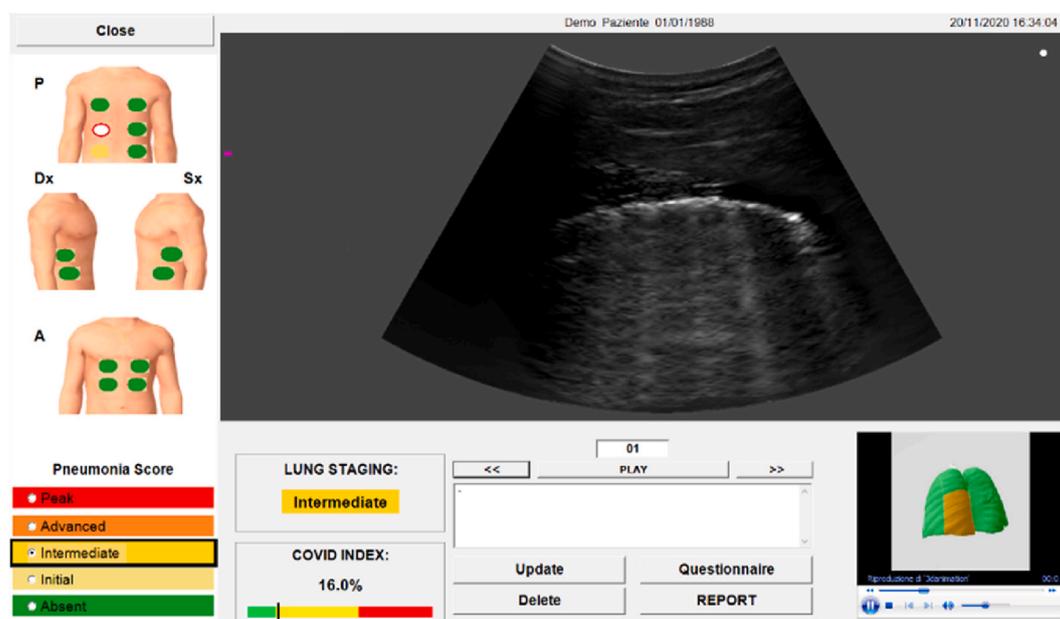


Fig. 1. Software interface after the acquisition. The image is a screenshot of the automatic data elaboration of the results, including Pneumonia Score, Lung Staging and COVID Index values. The main panel of the user interface shows the acquired B-mode images.

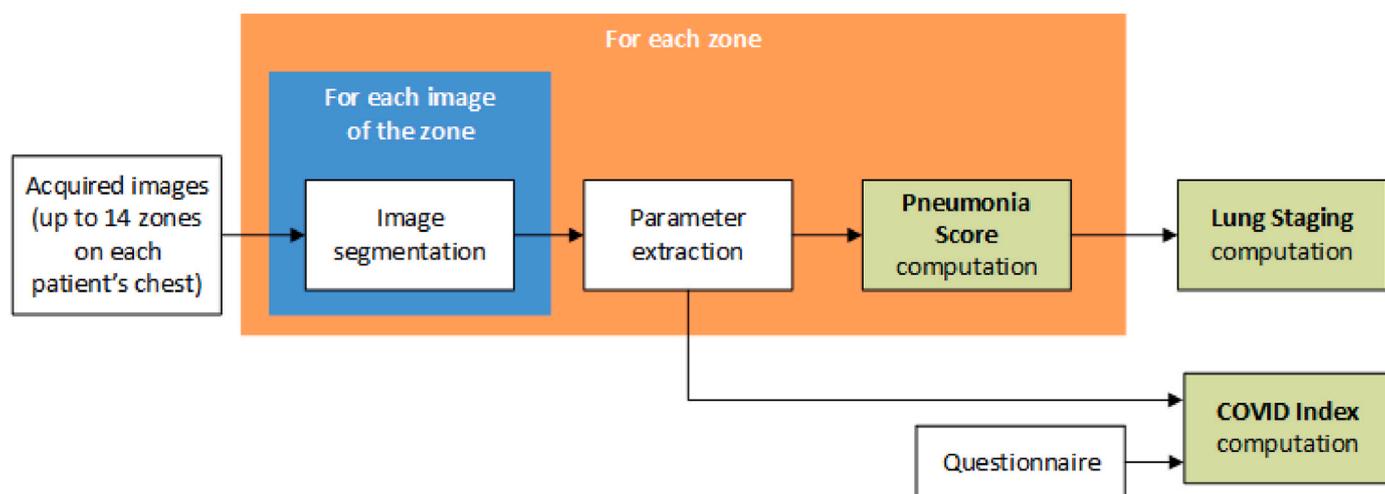


Fig. 2. Block Diagram: Working principle of the implemented algorithm.

All the images available for each considered patient, both longitudinal and transversal, were stored and off-line analysed by an expert sonographer, who was kept blind with respect to the results automatically provided by the software. The operator was asked to assign a *Pneumonia Score*, according to the scale explained above, to each scanned anatomical zone of each enrolled patient. The assessment performed by the expert operator was taken as reference and considered as “ground truth”. Similarly, the *Lung Staging* value obtained by the algorithm was compared to the value obtained by the expert operator considered as reference.

Finally, as regards the calculation of the probability that the pneumonia was due to SARS-CoV-2, the ultrasonographic data collected from patients for whom the positivity or negativity to the virus had been previously verified using an RT-PCR oropharyngeal swab were catalogued on the basis of the presence of specific ultrasound patterns, such as number and arrangements of the B-lines, consolidations, pleural modifications, etc. Thus, the software estimated the *COVID Index* (%) by integrating the presence of specific imaging patterns, usually presenting bilaterally and in basal lobes (patchy, no-gravity related B-lines, with very defined spared areas; sonographic *white lung*; irregular or fragmented pleural line; small peripheral consolidations or, rarely, larger consolidations in more advanced phases; trivial localized pleural effusion in the context of more deaerated areas [19]) with the clinical, demographic and epidemiological information derived from the questionnaire (as detailed in the “Measurements” paragraph) through a regression model. Each clinical and epidemiological information was associated to a specific score value, which had been previously determined through a statistical analysis of available data on the occurrence of the considered conditions in COVID-19 patients. The questionnaire result was then integrated with the image analysis result, calculated using a proprietary algorithm that acted in a similar way by assigning a specific score value to each detected pneumonia marker and to each further detected echographic sign of possible COVID-19 infection (being again based on the statistical analysis of available data on the presence of the considered echographic markers and signs in COVID-19 patients). The final *COVID Index* was obtained as a weighted average of the two partial results. Regarding the validation of the *COVID Index*, the predicted probability given by the *COVID Index* was correlated with the actual incidence of COVID-19 pneumonia in the considered patients (verified by RT-PCR).

2.5. Analysis

As concerning both the *Pneumonia Score* and the *Lung Staging*, the significance of the relationship between the 5-stage classification

performed by the algorithm and by the expert operator was assessed through the calculation of the Cohen’s K, the Chi-squared (χ^2) test and the Spearman’s correlation coefficient ρ . The accuracy of the algorithm in the *Pneumonia Score* and *Lung Staging* assignment with respect to the classification performed by the expert operator was evaluated as diagnostic concordance, *i.e.* the ability to give the same score as the expert operator, calculated as the ratio between the number of concordant scores and the overall number of assessed cases.

Moreover, in order to evaluate both the ability to distinguish between patients with or without pneumonia (*i.e.* *Lung Staging* > 0 or *Lung Staging* = 0, respectively) and the ability to distinguish between patients with mild and severe signs (*i.e.* *Lung Staging* ≤ 2 or *Lung Staging* > 2), the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR⁺) and the negative likelihood ratio (LR⁻) were also assessed [23].

The association between the *COVID Index* and the actual incidence of SARS-CoV-2 pneumonia was estimated considering 10-point ranges of *COVID Index* (*i.e.* groups of patients with *COVID Index* <10, ≥10 and < 20, ≥20 and < 30, etc.) and correlating the average *COVID Index* with the percentage of SARS-CoV-2 positive patients in each considered range. In principle, for each single *COVID Index* value the correlation with the percentage of patients actually infected by SARS-CoV-2 should be assessed. However, since a limited number of patients was available for each *COVID Index* value, the patients were grouped into 10-point ranges. In the plot reported in Fig. 6, coordinates of each point represent the mean *COVID Index* of the group of patients in the considered *COVID Index* range and the percentage of such patients that actually resulted positive to SARS-CoV-2. For example, the mean *COVID Index* of the patients in the range 40–50 was equal to 44.3% and 50.0% of them resulted positive to SARS-CoV-2. The strength of the correlation was assessed as Pearson correlation coefficient r and coefficient of determination r^2 . The accuracy of the *COVID Index* to discriminate between patients positive or negative to SARS-CoV-2 was investigated using Receiver Operating Characteristic (ROC) curve analysis [24,25].

We estimated that a sample size of 245 patients was necessary to observe a sensitivity and a specificity of 85%, considering a prevalence of the disease of 0.8 (intended as the proportion of the population arriving to the Emergency Department for suspected pneumonia that actually had the disease, estimated from a previous group of patients), with a type I error $\alpha = 0.05$ [26].

The statistical analysis was performed using MATLAB® R2018a (The MathWorks, Inc., Natick, MA) and MedCalc Statistical Software version 19.6 (MedCalc Software Ltd, Ostend, Belgium).

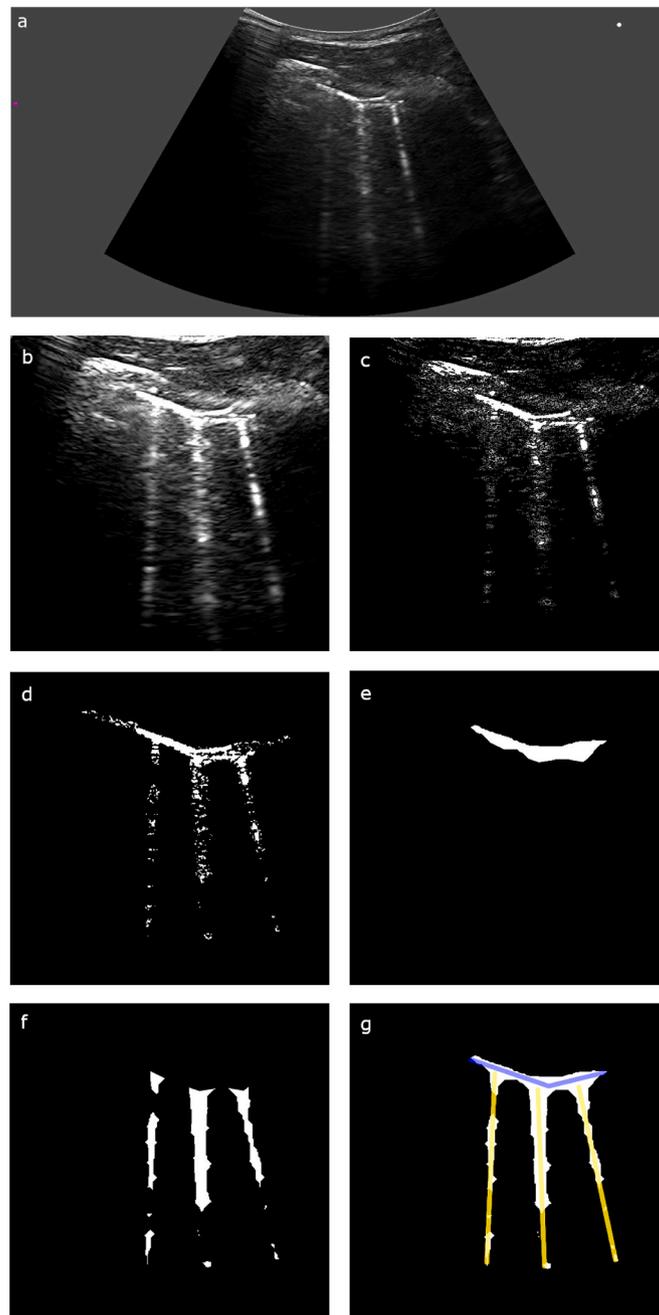


Fig. 3. Image segmentation steps applied to a typical transversal image: Application of image processing steps for automatic segmentation of an ultrasound frame containing pleura and three B-lines: (a) initial image; (b) image pre-processing; (c) 'raw' identification of target structures; (d) conversion to a binary map and median filter application; (e,f) hole-filling and pixel erosion/dilatation for the separate identification of pleura (e) and B-lines (f), taking also into account automatic selection criteria based on morphological considerations; (g) determination of dimensions and relative positions of the target structures.

3. Results

3.1. Characteristics of the study subjects

Overall, 556 cases were considered for this study. According to the expert operator evaluation, 124 patients (22.3%) were classified as *Lung Staging* = 0 (no disease), 184 (33.1%) as *Lung Staging* = 1 (early stage), 108 (19.4%) as *Lung Staging* = 2 (intermediate stage), 96 (17.3%) as *Lung Staging* = 3 (advanced stage) and 44 (7.9%) as *Lung Staging* = 4 (severe stage). The most commonly observed clinical manifestations of lung infections in the recruited patients were fever, hypoxemia, cough or other respiratory symptoms. The dataset was split into two datasets of

equal size (278 cases each, randomly assigned), which were independently used to implement and validate the proposed algorithm, respectively. Overall, 2421 zones were analysed.

3.2. Main results

Considering the independent validation dataset and assessing the performance of the algorithm in the *Pneumonia Score* assignment, the Spearman correlation coefficient was $\rho = 0.932$ (95% CI: 0.927 to 0.937, $p < 0.0001$), the χ^2 test was significant (contingency coefficient = 0.860, $p < 0.0001$) and Cohen's K was 0.901 (95% CI: 0.887 to 0.914). The overall rate of agreement between the *Pneumonia Score* assigned by the

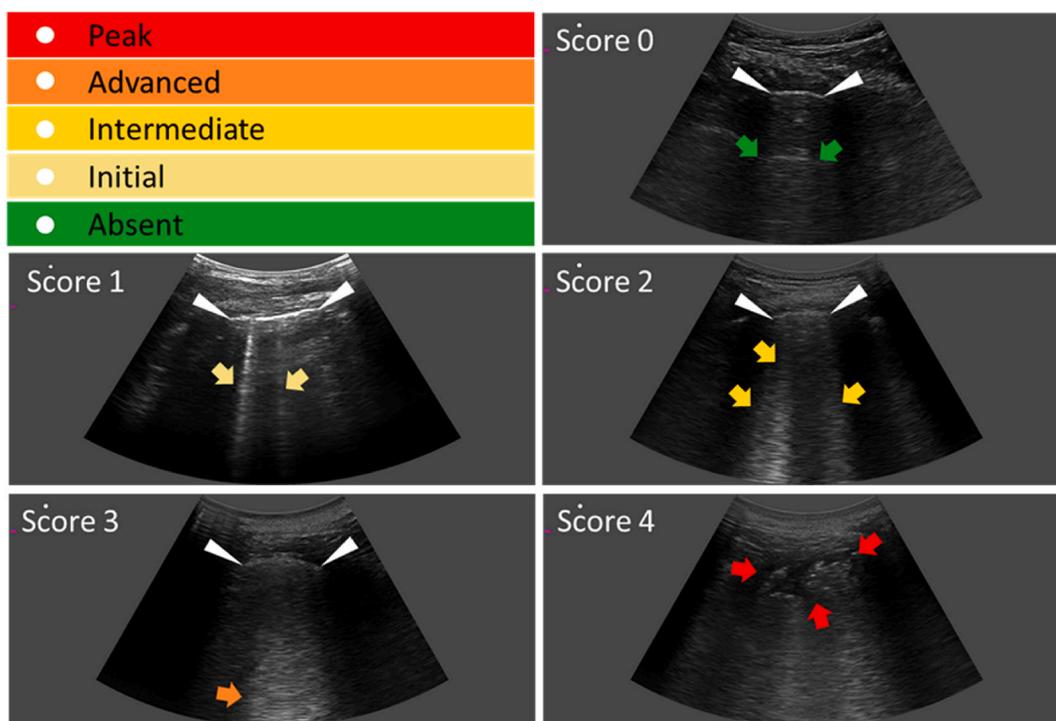


Fig. 4. Pneumonia score classification: Score 0 – Absent: a normal lung pattern was identified such as the presence of clear pleural line (highlighted by white triangular arrowheads) and A-line (highlighted by green arrows); Score 1 – Initial: a small number of B-lines, less than 3 (highlighted by pale yellow arrows); Score 2 – Intermediate: the presence of diffused and multifocal B-lines (highlighted by yellow arrows); Score 3 – Advanced: with diffused and multifocal B-lines and initial subpleural consolidations and white lung pattern (highlighted by the orange arrow); Score 4 – Peak: with lung consolidations (highlighted by red arrows). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

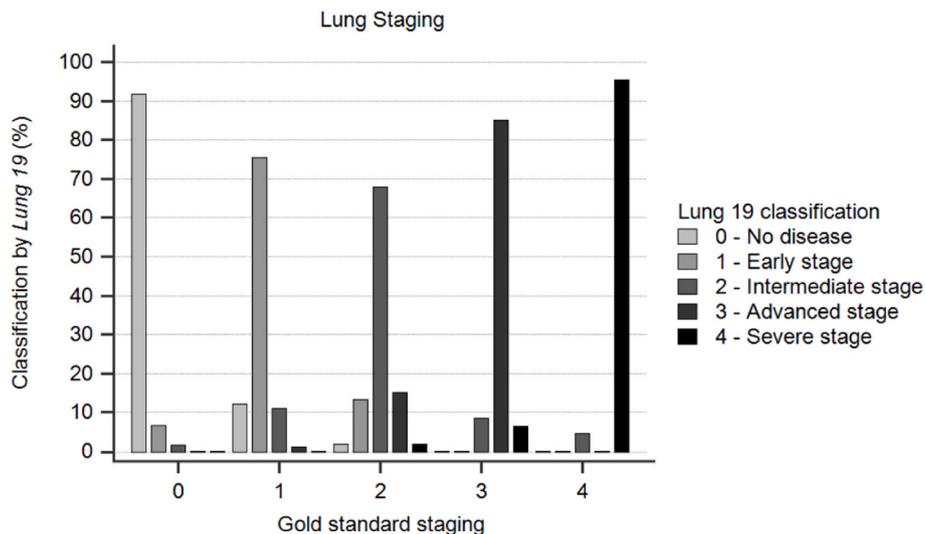


Fig. 5. – Histogram of the Lung Staging 5-class classification performed by the Lung 19 software with respect to the gold standard reference, i.e. the classification performed by the expert operator.

expert operator and the one assigned by the algorithm was 90.8%.

As considering the 5-grade *Lung Staging* classification, the Spearman correlation coefficient between the *Lung Staging* assigned by the automatic algorithm and the expert operator was $\rho = 0.916$ (95% CI: 0.894 to 0.933, $p < 0.0001$). At χ^2 test, a significant p-value was obtained ($p < 0.0001$), with a contingency coefficient of 0.843, showing that there was a strong relationship between the evaluation performed by the automatic algorithm and the expert operator. The distribution of classification among classes performed by the expert operator and the software algorithm is shown in Fig. 5. The Cohen’s K was 0.848 (95% CI: 0.808 to

0.888). The percentage of overall agreement in the 5-grade classification was 80.6%, meaning that in large majority of cases the diagnostic classification performed by the algorithm was in the same class as the one performed by the operator.

As concerning the capability to classify patients into sick and healthy (i.e. *Lung Staging* = 0 and *Lung Staging* > 0, respectively), the diagnostic agreement was 93.5%, the sensitivity was 94.4% and specificity was 90.3%. The PPV and NPV were 97.1% and 82.4%, respectively. As concerning the capability to classify patients with mild or severe signs, instead, (i.e. *Lung Staging* ≤ 2 and *Lung Staging* > 2, respectively), the

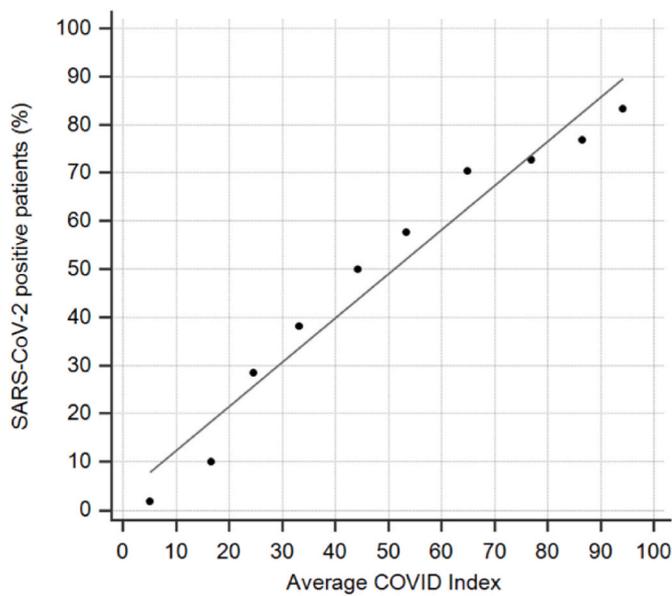


Fig. 6. Linear regression between average COVID Index per 10-point ranges and percentage of SARS-CoV-2 positive patients.

Table 1
2 by 2 used for the classification of healthy or sick patients.

		Expert operator	
		Disease Present	Disease Absent
automatic algorithm	Disease Present	204	6
	Disease Absent	12	56

Table 2
2 by 2 used for the classification of mild or severe patients.

		Expert operator	
		Mild Disease	Severe Disease
automatic algorithm	Mild Disease	124	5
	Severe Disease	11	64

Table 3
Algorithm performance evaluated with respect to the expert operator classification considered as reference, in the classification of healthy or sick patients and of patients with mild or severe signs.

STATISTIC (95% CI)	LUNG STAGING CLASSIFICATION IN HEALTHY AND SICK PATIENTS	LUNG STAGING CLASSIFICATION IN MILD AND SEVERE SYMPTOMS
DIAGNOSTIC CONCORDANCE	93.5% (90.0%–96.1%)	92.2% (87.6%–95.5%)
SENSITIVITY	94.4% (90.5%–97.1%)	92.0% (85.9%–95.9%)
SPECIFICITY	90.3% (80.1%–96.4%)	92.8% (83.9%–97.6%)
POSITIVE LIKELIHOOD RATIO (LR ⁺)	9.8 (4.6–20.9)	12.7 (5.4–29.5)
NEGATIVE LIKELIHOOD RATIO (LR ⁻)	0.06 (0.04–0.11)	0.09 (0.05–0.16)
POSITIVE PREDICTIVE VALUE (PPV)	97.1% (94.1%–98.6%)	96.1% (91.4%–98.3%)
NEGATIVE PREDICTIVE VALUE (NPV)	82.4% (72.8%–89.1%)	85.3% (76.7%–91.1%)
AREA UNDER THE CURVE (AUC)	0.948 (0.915–0.971)	0.954 (0.915–0.978)

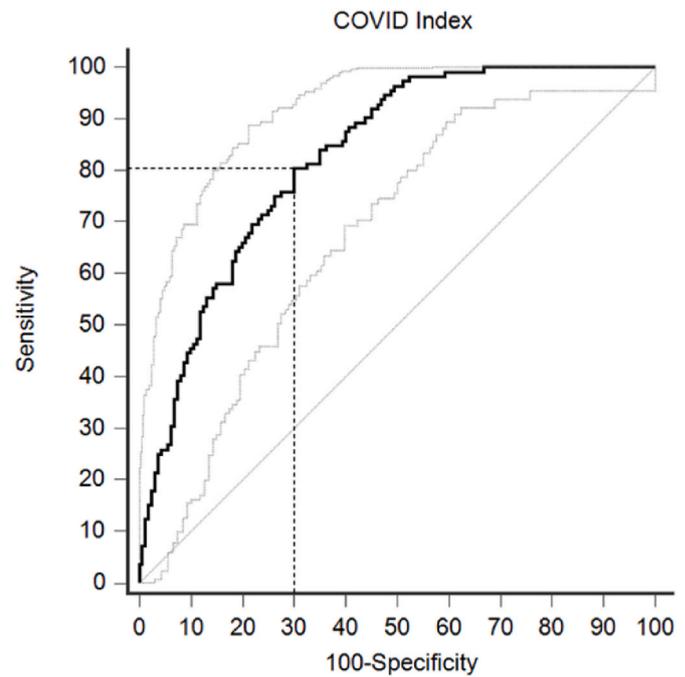


Fig. 7. ROC curve analysis of COVID Index ability to discriminate between patients positive or negative to COVID-19. The black bold line represents the ROC curve (AUC = 0.826) and the thin grey lines represent the 95% confidence interval. The dashed back lines indicate the value of sensitivity and specificity in correspondence of the Youden index.

diagnostic agreement was 92.2%, the sensitivity was 92.0% and specificity was 92.8%, whereas the PPV and NPV were 96.1% and 85.3%, respectively. The extensive results about the performance obtained with the validation dataset are reported in Tables 1–3.

Considering the prevalence of SARS-CoV-2 positivity compared to the average COVID index evaluated in steps of 10 units, the correlation was very high, with $r = 0.977$, $r^2 = 0.954$ and the slope of the regression line 0.92 (95% CI: 0.75 to 1.08). The regression distribution is reported in Fig. 6.

At ROC analysis (Fig. 7), the AUC was 0.826 (95% CI: 0.775 to 0.869, $p < 0.0001$). In correspondence of optimal cut-off associated to the Youden index, providing the best trade-off sensitivity and specificity is maximised, i.e. COVID Index > 35.5%, the sensitivity and specificity were 80.4% (95% CI: 71.8%–87.3%) and 70.0% (95% CI: 62.3%–77.0%), respectively.

4. Discussion

To meet the clinical needs related both to a standardization of an automatic quantitative pneumonia assessment and to the recently emerged SARS-CoV-2 pandemic, an algorithm based on a fully automatic ultrasonographic data processing for the identification of imaging patterns distinctive of pneumonia has been proposed and validated on a first set of patients. To the best of our knowledge, this is a unique approach for the automatic and quantitative analysis of the lung health status, giving a measure of the pneumonia severity and indicating the probability of its association with SARS-CoV-2. This technology might pave the way for a standardization of the pneumonia assessment using a radiation-free approach. Indeed, the results showed a high level of accuracy of the algorithm in distinguishing the cohort of healthy people from the sick patients with a very high level of sensitivity and, among the sick patients, the algorithm was able to correctly identify the lower stages (Lung Staging 1 or 2) with respect to the higher ones (Lung Staging 3 or 4). Overall, a substantial agreement between the 5-class staging performed by the algorithm and the gold standard method, i.e. the

staging performed by the expert operator, has been demonstrated. The majority of the discordances between the methods concern the early and intermediate stages, and only in 5 out of 278 cases (1.8%) a difference of 2 stages has been observed between the proposed algorithm and the expert operator.

The clinical performance obtained by the *Lung Staging* in the evaluation of the patient lungs reflected the even higher performance obtained by the *Pneumonia Score* in the assessment of each zone, separately considered. This might be due the intrinsic nature of *Lung Staging*, calculated as the maximum value of *Pneumonia Score* observed for a given patient: indeed, the high occurrence of 0 scores when the single zones are considered with respect to the overall lung evaluation, might contribute to the apparently better accuracy of the *Pneumonia Score* with respect to the *Lung Staging*. Moreover, when the *Lung Staging* is automatically calculated, a disagreement in a single *Pneumonia Score* might result in a *Lung Staging* disagreement, even if the scores were accurately assigned for the majority of analysed zones. This aspect underlines the usefulness of the possibility for the operator to confirm or to correct the automatically assigned *Pneumonia Score*.

The SARS-CoV-2 pandemic has led to an increasing request of point-of-care ultrasound in daily clinical practice as an extension to bedside clinical examination through the employment of small and portable devices [19]. It has been recently demonstrated that the LUS can be successfully employed to detect lung alterations attributable to SARS-CoV-2 infection, even earlier than the positivity assessed by RT-PCR, thus allowing to compensate the false negative results obtained by the oro/naso-pharyngeal swab [15]. However, being the results mainly derived from the individual interpretation of the acquired images, qualitative and subjective analyses might occur: in this context, the development of an automatic and quantitative analysis is urgently needed [27].

The two main limitations of this study are the lack of the assessment of the intra- and inter-operator variability in pneumonia staging using LUS imaging and the unavailability of clinical data to define severity.

Regarding the lack of variability assessment, with this initial settlement and validation of the software module, the variability in pneumonia staging was not considered crucial, also because there was no the actual possibility of re-scanning the patients due to the emergency context and to the use of some cases obtained from online databases: variability will be investigated with subsequent dedicated studies, involving smaller number of patients and just focused on the assessment of repeatability and reproducibility. This aspect is also connected with the involvement of different operators. On one side, this might be a disadvantage, since the evaluation of the performance might have been influenced by the contribution of the inter-operator variability, but on the other side the real-life use of LUS should be considered: though the use of this technology is simple and requires a relatively short training period [28], the employment of LUS has only recently spread in clinical practice. For this reason, the majority of the operators have a short experience. It would be of interest the assessment of the performance by varying the operator's experience, also to assess whether the guided acquisition and the automatic analysis offered by the software would be an aid in the reduction of the operator-related variability. In this context, the use of an expert operator analysis as a gold standard reference could be also considered a limitation of the study, although this is a very common approach in studies aimed to evaluate the accuracy of automatic or semi-automatic algorithms for ultrasound image processing for various purposes.

Regarding the absence of clinical data to define disease severity, actually, values such as oxygen saturation, oxygen support, PaO₂/FIO₂ ratio etc., were not available because the considered cases derived either from scanning performed in the emergency context or from online databases. However, the key object of the present study was to compare the automatic LUS score assigned by the algorithm with the score assigned by an expert operator based on the echographic image analysis only, and this should not be influenced by the mentioned clinical data.

Moreover, this study is a snapshot of a short-period activity during the SARS-CoV-2 pandemic, therefore characterized by the high prevalence of the COVID-19 pneumonia. The use of this technology in a standard clinical scenario should be investigated in subsequent studies.

Finally, the monitoring of the enrolled patients is missing in this report. Indeed, the addition of further analysis for the monitoring would have made it incompatible with the overloaded clinical routine, but it will be planned in further studies dedicated to the monitoring of pneumonia disease evolution.

As yet, there are no devices that can automatically evaluate ultrasonographic data derived from lung investigations, unless based on artificial intelligence approaches [29,30]. Nonetheless, important studies have been conducted before the onset of the SARS-CoV-2 pandemic, on the development of automatic or semi-automatic algorithms based on LUS image processing, but they are not commercially available so far. Interesting approaches for the automatic identification and quantitative assessment of B-lines in patients with pulmonary edema or acute respiratory distress have been recently proposed [10, 31], showing that the implementation of quantitative LUS through computer-aided scoring had potential benefits in terms of faster data analysis and applicability to large datasets at no additional cost. Interestingly, the method proposed in the present study is similar, though it integrates also the fully automatic analysis of ultrasonographic data and works in real-time. In a study by Corradi et al. [11], the diagnostic performance of a method for the quantitative analysis of LUS was compared with chest X-ray and visual assessment of LUS for the detection of pneumonia, using CT as the gold standard. Quantitative LUS showed greater sensitivity (93%), specificity (95%) and diagnostic accuracy (94%) than chest X-ray and visual ultrasound or their combination in diagnosing pneumonia. This is an evidence of how quantitative LUS might represent a valid support for the physicians dedicated to the triage phase and committed to quickly referring the patient to the appropriate care unit.

The methodology proposed in the present study allowed for the identification of patients suffering from pneumonia with high accuracy. The implemented method integrated the known advantages of the generic LUS (such as portability, non-invasiveness and real-time information), which allow for the use at bedside also in intensive care units, first aid departments and Special Continuity Care Units that provide home care for COVID-19 patients, with the automatic approach that increased the objectivity and gave a quantitative indicator of the pneumonia staging, with potentially important implications in terms of reduction of operator-related variability and usability regardless of the operator's experience.

5. Conclusions

This new LUS-based technology, described and preliminarily validated in this study, showed a good capability of diagnosing and staging pneumonia. The reported results unveil also the potential of the new LUS-based technology to be used to monitor pneumonia evolution in a personalized way (*i.e.*, by repeating the exam with the frequency deemed as appropriate by the physician). Moreover, in addition to a 5-stage pneumonia scoring given by the *Lung Staging* parameter, the statistical probability of being infected by SARS-CoV-2 was also expressed through the *COVID Index*, thus allowing the clinical staff to activate the most adequate clinical path.

This device might represent a significant decision support tool for both experienced and less experienced staff (in particular for personnel in training), due to the simplicity of its use, with a user-guided acquisition and a fully automatic real-time analysis. Indeed, it does not require the visual analysis of the acquired images, though it is given the operator the possibility to accept or to intervene to modify the automatically obtained result. The adoption of this approach within the triage and intensive care areas, as well as in the ambulance or at patient's home, might have important implications both at the population

level, since it would allow for mass screening and widespread prevention (particularly for the more fragile population, such as the elderly and patients with previous comorbidities or other risk factors), and at healthcare system level, offering a rapid and cost-effective identification of any-aetiology pneumonia in a simple and objective way.

CRediT authorship contribution statement

Fiorella Anna Lombardi: Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft. **Roberto Franchini:** Software, Formal analysis. **Rocco Morello:** Methodology, Software, Validation, Formal analysis, Investigation. **Ernesto Casciaro:** Data curation, Writing – review & editing. **Stefania Ianniello:** Conceptualization, Validation, Resources, Writing – review & editing. **Maurizio Serra:** Conceptualization, Validation, Resources, Supervision. **Francesco Satriano:** Conceptualization, Validation, Resources, Supervision. **Francesco Mojoli:** Writing – review & editing. **Silvia Mongodi:** Writing – review & editing. **Daniela Pignatelli:** Writing – review & editing. **Marco Di Paola:** Conceptualization, Validation, Data curation, Supervision. **Sergio Casciaro:** Conceptualization, Validation, Supervision, Project administration, Writing – review & editing.

Declaration of competing interest

The Authors declare that they have no conflicts of interest.

References

- [1] Y. Xia, Y. Ying, S. Wang, W. Li, H. Shen, Effectiveness of lung ultrasonography for diagnosis of pneumonia in adults: a systematic review and meta-analysis, *J. Thorac. Dis.* 8 (10) (2016 Oct) 2822–2831, <https://doi.org/10.21037/jtd.2016.09.38>.
- [2] M. Arabiat, A.E. Foderaro, A.T. Levinson, Lung ultrasound for diagnosing patients with severe dyspnea and acute hypoxic respiratory failure, *R. I. Med. J.* 102 (10) (2019 Dec 2) 34–38, 31795532.
- [3] L. Demi, T. Egan, M. Muller, Lung ultrasound imaging, a technical review, *Appl. Sci.* 10 (2) (2020) 462.
- [4] D.A. Lichtenstein, Lung ultrasound in the critically ill, *Ann. Intensive Care* 4 (2014) 1, <https://doi.org/10.1186/2110-5820-4-1>.
- [5] L. Long, H.T. Zhao, Z.Y. Zhang, G.Y. Wang, H.L. Zhao, Lung ultrasound for the diagnosis of pneumonia in adults: a meta-analysis, *Medicine (Baltim.)* 96 (3) (2017 Jan), e5713, <https://doi.org/10.1097/MD.00000000000005713>, 28099332, PMC5279077.
- [6] P.H. Mayo, R. Copetti, D. Feller-Kopman, G. Mathis, E. Maury, S. Mongodi, F. Mojoli, G. Volpicelli, M. Zanobetti, Thoracic ultrasonography: a narrative review, *Intensive Care Med.* 45 (9) (2019 Sep) 1200–1211, <https://doi.org/10.1007/s00134-019-05725-8>. Epub 2019 Aug 15,31418060.
- [7] A. Reissig, R. Copetti, Lung ultrasound in community-acquired pneumonia and in interstitial lung diseases, *Respiration* 87 (3) (2014) 179–189, <https://doi.org/10.1159/000357449>. Epub 2014 Jan 28,24481027.
- [8] D. Lichtenstein, G. Mézière, P. Biderman, A. Gepner, O. Barré, The comet-tail artifact. An ultrasound sign of alveolar-interstitial syndrome, *Am. J. Respir. Crit. Care Med.* 156 (5) (1997) 1640–1646, <https://doi.org/10.1164/ajrccm.156.5.96-07096>.
- [9] D. Lichtenstein, G. Mézière, A lung ultrasound sign allowing bedside distinction between pulmonary edema and COPD: the comet-tail artifact, *Intensive Care Med.* 24 (12) (1998) 1331–1334, <https://doi.org/10.1007/s001340050771>.
- [10] C. Brusasco, G. Santori, E. Bruzzo, R. Trò, C. Robba, G. Tavazzi, F. Guarracino, F. Forfori, P. Boccacci, F. Corradi, Quantitative lung ultrasonography: a putative new algorithm for automatic detection and quantification of B-lines, *Crit. Care* 23 (1) (2019 Aug 28) 288, <https://doi.org/10.1186/s13054-019-2569-4>, 31455421, PMC6712728.
- [11] F. Corradi, C. Brusasco, A. Garlaschi, F. Paparo, L. Ball, G. Santori, P. Pelosi, F. Altomonte, A. Vezzani, V. Brusasco, Quantitative analysis of lung ultrasonography for the detection of community-acquired pneumonia: a pilot study, *BioMed Res. Int.* 2015 (2015) 868707, <https://doi.org/10.1155/2015/868707>. Epub 2015 Feb 25,25811032,PMC4355628.
- [12] G. Via, E. Storti, G. Gulati, L. Neri, F. Mojoli, A. Braschi, Lung ultrasound in the ICU: from diagnostic instrument to respiratory monitoring tool, *Minerva Anestesiol.* 78 (11) (2012 Nov) 1282–1296. Epub 2012 Aug 3,22858877.
- [13] M. Allinovi, A. Parise, M. Giacalone, A. Amerio, M. Delsante, A. Odone, A. Franci, F. Gigliotti, S. Amadasi, D. Delmonte, N. Parri, A. Mangia, Lung ultrasound may support diagnosis and monitoring of COVID-19 pneumonia, *Ultrasound Med. Biol.* 46 (11) (2020 Nov) 2908–2917, <https://doi.org/10.1016/j.ultrasmedbio.2020.07.018>. Epub 2020 Jul 20,32807570,PMC7369598.
- [14] G. Giovannetti, L. De Michele, M. De Ceglie, P. Pierucci, A. Mirabile, M. Vita, V. O. Palmieri, G.E. Carpagnano, A. Scardapane, C. D'Agostino, Lung ultrasonography for long-term follow-up of COVID-19 survivors compared to chest CT scan, *Respir. Med.* 181 (2021 May) 106384, <https://doi.org/10.1016/j.rmed.2021.106384>. Epub 2021 Mar 31,33839587.
- [15] E. Pivetta, A. Goffi, M. Tizzani, S.M. Locatelli, G. Porrino, I. Losano, D. Leone, G. Calzolari, M. Vespa, F. Steri, A. Arditò, M. Capuano, M. Gelardi, G. Silvestri, S. Dutto, M. Avolio, R. Cavallo, A. Bartalucci, C. Paglieri, F. Morello, L. Richiardi, M.M. Maule, E. Lupia, Molinette MedUrg group on lung ultrasound. Lung ultrasonography for the diagnosis of SARS-CoV-2 pneumonia in the emergency department, *Ann. Emerg. Med.* 77 (4) (2021 Apr) 385–394, <https://doi.org/10.1016/j.annemergmed.2020.10.008>. Epub 2020 Oct 13. PMID: 33461884; PMID: PMC7552969.
- [16] F. Piscaglia, F. Stefanini, V. Cantisani, P.S. Sidhu, R. Barr, A. Berzigotti, M. C. Chammass, J.M. Correas, C.F. Dietrich, S. Feinstein, P. Huang, C. Jenssen, Y. Kono, M. Kudo, P. Liang, A. Lyshchik, C. Nolsoe, X. Xie, F. Tovoli, Benefits, Open questions and Challenges of the use of Ultrasound in the COVID-19 pandemic era. The views of a panel of worldwide international experts, *English, Ultraschall der Med.* 41 (3) (2020 Jun) 228–236, <https://doi.org/10.1055/a-1149-9872>. Epub 2020 Apr 15,32294795.
- [17] E. Urbankowska, K. Krenke, L. Drobaczyński, P. Korczyński, T. Urbankowski, M. Krawiec, G. Kraj, M. Brzewski, M. Kulus, Lung ultrasound in the diagnosis and monitoring of community acquired pneumonia in children, *Respir. Med.* 109 (9) (2015 Sep) 1207–1212, <https://doi.org/10.1016/j.rmed.2015.06.011>. Epub 2015 Jun 23,26138897.
- [18] M.D. Hope, C.A. Raptis, A. Shah, M.M. Hammer, T.S. Henry, six signatories, A role for CT in COVID-19? What data really tell us so far, *Lancet* 395 (10231) (2020 Apr 11) 1189–1190, [https://doi.org/10.1016/S0140-6736\(20\)30728-5](https://doi.org/10.1016/S0140-6736(20)30728-5). Epub 2020 Mar 27,32224299,PMC7195087.
- [19] L. Gargani, H. Soliman-Aboumarie, G. Volpicelli, F. Corradi, M.C. Pastore, M. Cameli, Why, when, and how to use lung ultrasound during the COVID-19 pandemic: enthusiasm and caution, *Eur Heart J Cardiovasc Imaging* 21 (9) (2020 Sep 1) 941–948, <https://doi.org/10.1093/ehjci/jeaa163>, 32515793,PMC7314093.
- [20] World Health Organization, Population-Based Age-Stratified Seroepidemiological Investigation Protocol for COVID-19 Virus Infection, 17 March 2020, World Health Organization, 2020. <https://apps.who.int/iris/handle/10665/331656>.
- [21] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D.F. Briganti, S. Perlini, E. Torri, A. Mariani, E.E. Mossolani, F. Tursi, F. Mento, L. Demi, Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method, *J. Ultrasound Med.* 39 (7) (2020 Jul) 1413–1419, <https://doi.org/10.1002/jum.15285>. Epub 2020 Apr 13,3227492,PMC7228287.
- [22] A. Smargiassi, G. Soldati, E. Torri, F. Mento, D. Milardi, P. Del Giacomo, G. De Matteis, M.L. Burzo, A.R. Larici, M. Pompili, L. Demi, R. Inchingolo, Lung ultrasound for COVID-19 patchy pneumonia: extended or limited evaluations? *J. Ultrasound Med.* 40 (3) (2021 Mar) 521–528, <https://doi.org/10.1002/jum.15428>. Epub 2020 Aug 20,32815618.
- [23] I.A. Gardner, M. Greiner, Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests, *Vet. Clin. Pathol.* 35 (1) (2006) 8–17, <https://doi.org/10.1111/j.1939-165x.2006.tb00802.x>.
- [24] C.E. Metz, Basic principles of ROC analysis, *Semin. Nucl. Med.* 8 (4) (1978 Oct) 283–298, [https://doi.org/10.1016/s0001-2998\(78\)80014-2](https://doi.org/10.1016/s0001-2998(78)80014-2), 112681.
- [25] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clin. Chem.* 39 (4) (1993 Apr) 561–577. Erratum in: *Clin Chem* 1993 Aug;39(8):1589,8472349.
- [26] N.M. Buderer, Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity, *Acad. Emerg. Med.* 3 (9) (1996 Sep) 895–900, <https://doi.org/10.1111/j.1553-2712.1996.tb03538.x>, 8870764.
- [27] P.I. Pietersen, K.R. Madsen, O. Graumann, L. Konge, B.U. Nielsen, C.B. Laursen, Lung ultrasound training: a systematic review of published literature in clinical lung ultrasound training, *Crit. Ultrasound J.* 10 (1) (2018 Sep 3) 23, <https://doi.org/10.1186/s13089-018-0103-6>, 30175392,PMC6119680.
- [28] L. Demi, Lung ultrasound: the future ahead and the lessons learned from COVID-19, *J. Acoust. Soc. Am.* 148 (4) (2020 Oct) 2146, <https://doi.org/10.1121/10.0002183>, 33138522.
- [29] C. Baloescu, G. Toporek, S. Kim, K. McNamara, R. Liu, M.M. Shaw, R.L. McNamara, B.I. Raju, C.L. Moore, Automated lung ultrasound B-line assessment using a deep learning algorithm, *IEEE Trans. Ultrason. Ferroelectrics Freq. Control* 67 (11) (2020 Nov) 2312–2320, <https://doi.org/10.1109/TUFFC.2020.3002249>. Epub 2020 Jun 15,32746183.
- [30] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli, E. Peschiera, R. Trevisan, G. Maschietto, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, P. Rota, A. Passerini, R.J.G. van Sloun, E. Ricci, L. Demi, Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound, *IEEE Trans. Med. Imag.* 39 (8) (2020 Aug) 2676–2687, <https://doi.org/10.1109/TMI.2020.2994459>. Epub 2020 May 14,32406829.
- [31] R. Moshavegh, K.L. Hansen, H. Moller-Sorensen, M.B. Nielsen, J.A. Jensen, Automatic detection of B-lines in in vivo lung ultrasound, *IEEE Trans. Ultrason. Ferroelectrics Freq. Control* 66 (2) (2019 Feb) 309–317, <https://doi.org/10.1109/TUFFC.2018.2885955>. Epub 2018 Dec 10,30530325.