



EDITORIAL

Open Access



To B or not to B-lines

Filipe André Gonzalez^{1,2,3*}, Jacobo Bacariza², Joao Leote^{2,4} and on behalf of EchoCrit Group

With great interest, we have read the paper from Boero et al. titled “Lung ultrasound among Expert operators: Scoring and Inter-rater Reliability Analysis (LESSON study), a secondary COWS study analysis from the ITA-LUS group” [1], which provides a focused evaluation of lung ultrasound (LUS). The author’s analysis reflects the increasing use of LUS as a valuable diagnostic and monitoring tool for assessing pulmonary conditions worldwide [2, 3]. The study is particularly noteworthy for emphasizing the inter-rater reliability of the LUS score among expert practitioners, whose assessments are crucial in determining the relevance of LUS in clinical practice [4]. By analyzing data from skilled clinicians, the article highlights the reliability of the LUS score when labeling ultrasound (US) video clips recorded from patients with COVID-19 pneumonia [5]. This focus is valuable, as skilled performance can establish benchmarks to define training standards for clinical practice through LUS assessment standardization.

Boero and colleagues [1] contribute to the expanding body of literature emphasizing the diagnostic precision of LUS, particularly when performed by trained professionals. Their findings align with recent analyses by other authors, confirming the moderate to high agreement rates achievable with LUS [6] and supporting its role in diagnosing acute lung conditions. One limitation of the

study is its focus on practitioners at the expert level. While this provides valuable insights into best practices, it restricts the generalizability of the findings to a broader range of clinicians at novice and intermediate levels. This limitation is significant given that LUS is frequently utilized as a frontline diagnostic tool in diverse settings [7]. Future research should aim to broaden the participant pool to include operators with varying levels of experience. This approach would afford a more comprehensive understanding of training requirements across different clinical environments. As in cardiac ultrasound settings, Gonzalez et al. and Varudo et al. [8, 9] already showed that machine learning-enabled real-time measurements of left ventricle ejection fraction and left ventricle out-flow tract velocity–time integral were strongly correlated with manual measurements, and the reproducibility was better with the machine learning system, including for novices. Besides the standardized training and interpretation that could minimize inter-operator variability among novices, using artificial intelligence (AI) tools for LUS could further improve inter-operator variability and allow less experienced users to use LUS more liberally [10].

Regarding the US system used for imaging recording, the authors used curvilinear probes with different frequencies (2–9 MHz) and different machines, including either conventional systems or ultraportable systems (i.e., Butterfly iQ) [1, 5]. In addition, the authors selected and gave, for experts’ evaluation, a proportional number of video clips recorded in each system to avoid a US system bias. However, this methodological option also promoted US imaging variability due to different lung assessment presets. Leote et al. evaluated the influence of US imaging settings on vertical artifacts (VA, used to mimic B-lines) in two phases. First, an *in vitro* phantom model demonstrated that variation of most of the US parameters

*Correspondence:

Filipe André Gonzalez

filipeandregonzalez@edu.ulisboa.pt; filipeandregonzalez@gmail.com

¹ Cardiovascular Research Center, Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal

² Intensive Care Department, Hospital Garcia de Orta EPE, Almada, Portugal

³ ICU in Hospital CUF Tejo, Lisbon, Portugal

⁴ Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Lisbon, Portugal



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

did not significantly affect the number and scoring of optimal VA [11]. Even though artifact intensity correlated strongly with power, gain, frequency, and dynamic range, the latter increased the number of discernible VA to 3 (from 36 to 102 dB). Second, an in vivo study on 29 patients under passive invasive mechanical ventilation showed a mild influence on the VA number after controlling for physiological and operator LUS confounders [12]. As in in vitro phantoms, the dynamic range also significantly increased the VA number recognized on invasively ventilated patients (Fig. 1, note the VA identified with asterisks after increasing from 60 to 102 dB). The authors concluded that to avoid a negative impact of US settings on VA, the system preset should use a lower probe frequency (i.e., 2 to 4 MHz), with a gain and power adjusted for a value near the available upper limit (i.e., gain 90%, power -10 dB) with a hyperechoic pleura (avoiding excessive brightness), an intermediate value of dynamic range adjusted for a discernable background media contrast (i.e., between 60 and 84 dB), without any post-processing tool such as artifact reduction, speckle reduction, frame averaging, or image enhancement. Their findings are supported by other report in the literature [13]. Regarding imaging interpretation, Boero et al. [1] variations in standard deviation results observed among clinicians were noted when distinguishing LUS score grades 0 and 1 (representing 0 to 2 B-lines or B-lines occupying less than 50% of the pleura, respectively). As described by other authors [6, 12], imaging interpretation when numbering VA is prone to some degree of inter-rater variability, even when employing a motionless probe and a

controlled inspiratory volume, across three clinicians [12].

LUS is used for diagnosis and treatment guidance, with repeated measures on the clinical evolution of a patient. If we add to the 20% inter-rater reliability in experienced users (on the LESSON study), more 15–30% for the variability of US settings (i.e., detecting 1 or 2 instead of 3 to 5 B-lines), and eventually more 10–20% among inexperienced users, we could easily reach an unacceptable inter-rater reliability. Hence, the forthcoming studies within the LUS community need to incorporate considerations for US settings [2, 7, 11–14].

As pointed by Boero and colleagues [1], LUS in clinical practice is based on a qualitative and subjective evaluation. Nonetheless, recent developments suggest using quantitative approaches to estimate the alveolar geometry after varying the probe's center frequency [15, 16]. While the LESSON study presents a structured approach to scoring, it would benefit from a more detailed explanation of the scoring criteria. Gonzalez et al. proposed a machine learning (ML) framework (using dimensionality reduction) for automated severity analysis of COVID-19 lung ultrasounds, intended to detect frames presenting alterations indicative of the disease and provide its subsequent severity classification [17]. These results showed that the empirically used grades for LUS scores had the corresponding clinical preponderance. For example, having more B-lines in the upper lung regions in the context of a cardiogenic pulmonary edema is much more severe, as there is a cephalocaudal gradient in the edema formation, usually starting in the basal lung regions. Moreover,

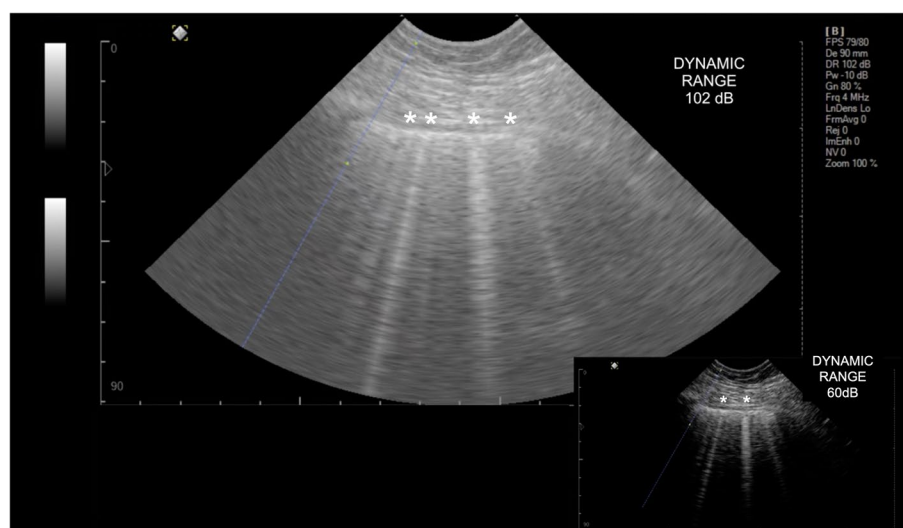


Fig. 1 Influence of dynamic range on the number of B-lines identified (asterisk). In the bigger lung ultrasound image, more B-lines can be detected when using a dynamic range of 102 dB, compared to the smaller image at the inferior right corner, which was recorded using a dynamic range of 60 dB (four vs two B-lines)

in some lung conditions, such as COVID-19 pneumonia, subpleural consolidations in the upper regions were associated with cardiac dysfunction and invasive mechanical ventilation [18]. Additionally, ML (support vector machine) models based on LUS were explored to improve decisions on ICU admission [19]. Using total values of LUS scores seems to result in the loss of valuable information. In fact, by leveraging ML with individual LUS findings, the prediction of ICU admission seems to be improved.

Implications and future directions

The findings of the LESSON study have significant implications for clinical practice. By confirming the reliability of LUS among expert practitioners, the study supports integrating LUS into standard assessment protocols for pulmonary conditions. However, the LUS should move into a more precise and accurate assessment. To hit the marks, it is mandatory to decrease the inter-observer variability by standardization of LUS in clinical practice, controlling the technical parameters for patient assessment, such as type of probe, US imaging settings, and evaluating lung regions first [1, 2, 4, 14]. Also, the upcoming research should encompass a broader range of operators and address automated VA detection algorithms based on artificial intelligence-enhanced diagnostics, which may further enhance precision [20, 21].

In summary, we suggest a three-step approach for LUS standardization in clinical practice:

1. Update the current guidelines [2, 3, 14], creating a practical guide on the use of LUS and B-lines interpretation, considering the following:
 - a. Recommending the most appropriate US settings for each probe and the main lung disease (i.e., pneumonia, edema, fibrosis). For example, according to our studies, the most fitted US settings for B-lines detection using a curvilinear probe are as follows: mechanical index of 0.5, depth range between 10 and 12 cm (at least 6 cm below the pleura), one focal depth (on the pleura) the power of -10 dB, gain of 90%, and equal level of time gain compensation across depth, standard line density, a dynamic range of 60 to 84 dB, center frequency of 4 MHz, without tissue equalization, or optional post-processing tools.
 - b. Choosing the best probe for each lung disease: a detailed image of the pleura may be needed for lung sliding or small consolidations, whereas for depth lung parenchyma, a low frequency should be selected.

- c. Concise reports and inherent conclusions should also be normalized to increase the applicability of LUS across various clinical settings.
 - d. Defining a consensual LUS score, either global and integrating regional preponderances to grade severity (with an online tool or app to calculate it easily).
2. LUS training and interpretation recommendations to ensure basic skills and keep an acceptable inter-rater reliability (e.g., 20% as in the LESSON study):
 - a. A checklist of minimal training competencies for different levels of expertise.
 - b. List of defined specialized reference centers for training.
3. Integrating LUS AI tools where possible to standardize acquisition, measurement, and interpretation for faster and more user-friendly daily use, enabling less trained clinicians to participate.

Acknowledgements

The EchoCrit Group is a collaboration group of the Intensive Care Department of Hospital Garcia de Orta, Almada, Portugal, dedicated to the advanced echocardiography and POCUS in critical care, formed by the following physicians: Filipe Gonzalez (consortia representative)¹, Rui Gomes¹, Jacobo Bacariza¹, Rita Varudo¹, João Leote¹, Vera Pereira¹, Dário Batista¹, Vânia Brito¹, Corinna Lohmann¹, João Gouveia¹, Joana Manuel¹, Liliana Santos¹, Sara Lança¹, Lucinda Oliveira¹, Tiago Ferreira¹, Joana Ferreira¹, João Sampaio¹, José Seoane¹, Inês Pimenta¹, Cristina Martins¹, Ricardo Meireles¹, Francisco D'Orey¹, Maria Inês Ribeiro¹, and Antero Fernandes¹ (Head of Intensive Care Department). The authors would like the names of the individual members of the group to be searchable through their individual PubMed records.

¹Intensive Care Department, Hospital Garcia de Orta, Almada, Portugal

Authors' contributions

FAG, JB and JL wrote the manuscript. The EchoCrit Group revised the manuscript. All authors revised the manuscript.

Funding

Not applicable.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 20 August 2024 Accepted: 26 August 2024

Published online: 05 September 2024

References

- Boero E, Gargani L, Schreiber A, Rovida S, Martinelli G, Maggiore SM, Urso F, Camporesi A, Tullio A, Lombardi FA et al (2024) Lung ultrasound among expert Operator'S: ScOring and INter-Rater Reliability Analysis (LESSON study) a secondary COWS study analysis from ITALUS group. *J Anesth Analg Crit Care* 4:1–9. <https://doi.org/10.1186/S44158-024-00187-X/FIGURES/4>
- Demi L, Wolfram F, Klersy C, De Silvestri A, Ferretti VV, Muller M, Miller D, Feletti F, Welnicki M, Buda N et al (2023) New international guidelines and consensus on the use of lung ultrasound. *J Ultrasound Med* 42:309. <https://doi.org/10.1002/JUM.16088>
- Volpicelli G, Elbarbary M, Blaivas M, Lichtenstein DA, Mathis G, Kirkpatrick AW, Melniker L, Gargani L, Noble VE, Via G et al (2012) International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med* 38:577–591. <https://doi.org/10.1007/S00134-012-2513-4>
- Vetruigno L, Biasucci DG, Deana C, Spadaro S, Lombardi FA, Longhini F, Pisani L, Boero E, Cereser L, Cammarota G et al (2024) Lung ultrasound and supine chest X-ray use in modern adult intensive care: mapping 30 years of advancement (1993–2023). *Ultrasound J* 16:1–12. <https://doi.org/10.1186/S13089-023-00351-4/FIGURES/3>
- Boero E, Rovida S, Schreiber A, Berchiolla P, Charrier L, Cravino MM, Converso M, Gollini P, Puppo M, Gravina A et al (2021) The COVID-19 Worsening Score (COWS)—a predictive bedside tool for critical illness. *Echocardiography* 38:207–216. <https://doi.org/10.1111/ECHO.14962>
- Anderson KL, Fields JM, Panebianco NL, Jenq KY, Marin J, Dean AJ (2013) Inter-rater reliability of quantifying pleural B-lines using multiple counting methods. *J Ultrasound Med* 32:115–120. <https://doi.org/10.7863/JUM.2013.32.1.115>
- Kamilaris A, Kramer JA, Baraniecki-Zwil G, Shofer F, Moore C, Panebianco N, Chan W (2023) Development of a novel observed structured clinical exam to assess clinical ultrasound proficiency in undergraduate medical education. *Ultrasound J* 15:1–8. <https://doi.org/10.1186/S13089-023-00337-2/FIGURES/4>
- Gonzalez FA, Varudo R, Leote J, Martins C, Bacariza J, Fernandes A, Michard F (2022) Automation of sub-aortic velocity time integral measurements by transthoracic echocardiography: clinical evaluation of an artificial intelligence-enabled tool in critically ill patients. *Br J Anaesth* 129:e116–e119. <https://doi.org/10.1016/j.bja.2022.07.037>
- Varudo R, Gonzalez FA, Leote J, Martins C, Bacariza J, Fernandes A, Michard F (2022) Machine learning for the real-time assessment of left ventricular ejection fraction in critically ill patients: a bedside evaluation by novices and experts in echocardiography. *Crit Care* 26:386. <https://doi.org/10.1186/s13054-022-04269-6>
- Russell, F.M.; Ehrman, R.R.; Barton, A.; Sarmiento, E.; Ottenhoff, J.E.; Nti, B.K. B-line quantification: comparing learners novice to lung ultrasound assisted by machine artificial intelligence technology to expert review. *Ultrasound J*. 2021;13. <https://doi.org/10.1186/S13089-021-00234-6>
- Leote J, Muxagata T, Guerreiro D, Francisco C, Dias H, Loução R, Bacariza J, Gonzalez F (2023) Influence of ultrasound settings on laboratory vertical artifacts. *Ultrasound Med Biol* 49:1901–1908. <https://doi.org/10.1016/J.ULTRASMEDBIO.2023.03.018>
- Leote, J.; Gonçalves, A.; Fonseca, J.; Loução, R.; Dias, H.; Ribeiro, M.I.; Meireles, R.; Varudo, R.; Bacariza, J.; Gonzalez, F. Impact of ultrasound settings on lung vertical artifacts: an observational study in mechanically ventilated patients. *ERJ Open Res*. 2024;00483–02024. <https://doi.org/10.1183/23120541.00483-2024>
- Duggan NM, Goldsmith AJ, Saud AAA, Ma IWY, Shokoohi H, Liteplo AS (2022) Optimizing lung ultrasound: the effect of depth, gain and focal position on sonographic B-lines. *Ultrasound Med Biol* 48:1509–1517. <https://doi.org/10.1016/j.ultrasmedbio.2022.03.015>
- Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, Perlini S, Torri E, Mariani A, Mossolani EE et al (2020) Proposal for international standardization of the use of lung ultrasound for patients with COVID-19. *J Ultrasound Med* 39:1413–1419. <https://doi.org/10.1002/JUM.15285>
- Leote J, Loução R, Aguiar M, Tavares M, Ferreira P, Muxagata T, Guerreiro D, Dias H, Bacariza J, Gonzalez F (2024) Total signal intensity of ultrasound laboratory vertical artifacts: a semi-quantitative tool. *WFUMB Ultrasound Open* 2:100035. <https://doi.org/10.1016/J.WFUMB.2024.100035>
- Mento F, Khan U, Fata F, Smargiassi A, Inchingolo R, Perrone T, Demi L (2022) State of the art in lung ultrasound, shifting from qualitative to quantitative analyses. *Ultrasound Med Biol* 48:2398. <https://doi.org/10.1016/J.ULTRASMEDBIO.2022.07.007>
- Gonzalez FA, Leote J, Sequeira M, Varudo R, Bacariza J, Gomes R, Meireles R, Martins C, Ribeiro J, Krippahl L, Bispo R, Fernandes A. Deep learning in COVID-19 LUS analysis—what can we use for the future? 1Machine learning improves ICU admission based on lung ultrasound score. *ESICM LIVES 2023*. *Intensive Care Med Exp*. 2023;11: 1–655. <https://doi.org/10.1186/S40635-023-00546-Y>
- Leote J, Judas T, Broa AL, Lopes M, Abecasis F, Pintassilgo I, Gonçalves A, Gonzalez F (2022) Time course of lung ultrasound findings in patients with COVID-19 pneumonia and cardiac dysfunction. *Ultrasound J* 14:1–11. <https://doi.org/10.1186/S13089-022-00278-2/FIGURES/3>
- Oliveira-Saraiva, D., Leote, J., Garcia, N., Gonzalez, F.A., (2024) Machine learning improves ICU admission based on lung ultrasound score. 43rd International Symposium on Intensive Care & Emergency Medicine. *Critical care*. 28 (S1),68. <https://doi.org/10.1186/s13054-024-04822-5>
- Oliveira-Saraiva, D.; Mendes, J.; Leote, J.; Gonzalez, F.A.; Garcia, N.; Ferreira, H.A.; Matela, N. Make it less complex: autoencoder for speckle noise removal—application to breast and lung ultrasound. *Journal of Imaging* 2023, Vol. 9, Page 217 2023, 9, 217, <https://doi.org/10.3390/JIMAGING9100217>.
- Shokoohi H, Lesaux MA, Roohani YH, Liteplo A, Huang C, Blaivas M (2019) Enhanced point-of-care ultrasound applications by integrating automated feature-learning systems using deep learning. *J Ultrasound Med* 38:1887–1897. <https://doi.org/10.1002/JUM.14860>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.