ORIGINAL RESEARCH PAPERS

Quantitative lung ultrasound spectroscopy: First comparison with gold standard computed tomography scan and standard lung ultrasound for diagnosis of pneumonia versus cardiogenic pulmonary edema

MATTIA PERPENTI[1*], ELEONORA BALZANI[1,2*], FEDERICO MENTO[1], CLAUDIA MARINARO[3], GIACOMO BELLANI[2], TIZIANO PERRONE[3], LIBERTARIO DEMI[1]

[1]Department of Information and Communication Engineering, University of Trento, Trento, Italy; [2]Interdepartmental Center of Medical Science, University of Trento, Trento, Italy; [3]Department of Internal Medicine, Humanitas Gavazzeni Bergamo, Bergamo, Italy. [*]These authors contributed equally to this work.

ABSTRACT

Background: Pneumonia (PNE) and cardiogenic pulmonary edema (CPE) are characterized by reduced airspaces dimension and edema. Their distinction through gold standard computed tomography (CT) is challenging due to their pattern similarity. Lung Ultrasound (LUS) is a tool for monitoring the progression of lung pathologies. LUS is portable, real-time, and non-ionized; however, standard LUS (S-LUS) relies on the subjective visualization of imaging patterns, leading to poor reproducibility and lack of diagnostic specificity. To enhance LUS diagnostic utility, quantitative LUS (Q-LUS) was developed. Q-LUS quantifies imaging patterns and explores their correlation to different pathophysiological conditions. In literature, vertical artifacts (VA) quantification proved capable of differentiating PNE and CPE, however, this approach was never compared with gold standard.

Methods: We statistically investigate and compare the clinical significance of CT, S-LUS, and Q-LUS, in differentiating PNE and CPE. From a cohort of 55 patients, CT, S-LUS, and Q-LUS data are acquired. CT and S-LUS data of each patient are evaluated to assign a semi-quantitative CT-score and S-LUS-score. Q-LUS radiofrequency data are acquired in multifrequency with convex (2, 3, and 4 MHz) and linear (3, 4, 5, and 6 MHz) probes. VA are

Received: 28 January 2026 | Accepted: 22 March 2025

Correspondence: Tiziano Perrone / Department of Internal Medicine, Humanitas Gavazzeni Bergamo, Via Mauro Gavazzeni, 21, Bergamo, 24125, Italy / E-mail: tiziano.perrone@gavazzeni.it Libertario Demi, Department of Information and Communication Engineering, University of Trento, Via Sommarive, 9, Trento, 38123, Italy / E-mail: libertario.demi@unitn.it

manually segmented, quantified into three spectral quantities, and statistically analyzed to extract 15 features for each patient. The diagnostic significance of the scores is tested through Generalized Estimating Equation models. **Results and Conclusions:** Results show areas under the curve of 74%, 72%, 65%, and 53% for Q-LUS linear, Q-LUS convex, CT-score, and S-LUS-score, respectively, highlighting Q-LUS as the most significant tool.

Key words: Cardiogenic pulmonary edema, computed tomography, lung ultrasound, pneumonia, quantitative lung ultrasound

Introduction

There are clinical scenarios in the branch of respiratory medicine, where the diagnosis of lung diseases is hindered due to similarities across histopathologies [i.e., pneumonia (PNE) and cardiogenic pulmonary edema (CPE)]. However, the underlying pathophysiological mechanisms inducing to a pathological condition could be different (e.g., the presence of pulmonary edema) or coexist within the same patient. PNE can lead to non-cardiogenic pulmonary edema by increasing alveolar-capillary permeability through local inflammation. This results in the accumulation of protein-rich fluid inside the alveoli due to widespread epithelial and endothelial injury (1). Several cytokines, including IL-6, $TNF-\alpha$, and $IL-1\beta$, are known to disrupt the integrity of the alveolar-capillary barrier, allowing albumin and other plasma proteins to leak into the lung parenchyma at high concentration (> 70% of the plasma content). The higher the cytokine levels, the more pronounced the leakage across the barrier becomes (2, 3). CPE, on the other hand, arises through a very different mechanism. Here, fluid shifts occur primarily due to elevated hydrostatic pressure, typically from left ventricular dysfunction, while the alveolar-capillary membrane remains largely intact. This results in the movement of a more protein-poor fluid, usually containing less than 50% of the plasma protein content (4). These different pathophysiological mechanisms are clinically relevant, as they guide distinct therapeutic strategies: ventilatory and anti-inflammatory support for PNE, and diuretics and preload reduction for CPE. While their diagnosis is essentially clinical, overlapping scenarios may occur, as observed during COVID-19, in which differentiating between PNE and CPE is crucial to guide appropriate therapy. Gold standard CT scan, while highly sensitive, lack specificity in differentiating PNE and CPE, underscoring a persistent diagnostic challenge (5).

Indeed, CT provides an optimal visualization of subtle or multilobar opacities and complications such as abscess or empyema, which can influence management in complex or immunocompromised patients. However, CT specificity is inconsistent (7–100%) due to overlap with other pulmonary diseases. Additionally, compared to other imaging modalities, CT entails high radiation exposure (1–8 mSv), high costs ($150–$1,500), and is poorly transportable. Lung ultrasound (LUS) has become increasingly popular at the bedside for evaluating respiratory failure. Indeed, this tool is capable of real-time imaging and is characterized by non-ionizing radiation exposure. Standard LUS (S-LUS) in medical practice relies on the subjective interpretation of imaging patterns, mostly the vertical artifacts (VA) for their association to pathological conditions. Indeed, although VA genesis remains unsolved, their visualization seems associated with the presence of acoustic traps along the lung surface (6). Acoustic traps are localized areas on the lung surface presenting a reduction of the air-spaces dimension. When an ultrasound wave at a specific frequency enters into a trap, it generates multiple scattering phenomena that are repeatedly radiated back to the ultrasound probe. This phenomenon generates a VA in the image that starts from the pleural line and moves synchronously with the pleural sliding. In S-LUS, the presence of VA is associated

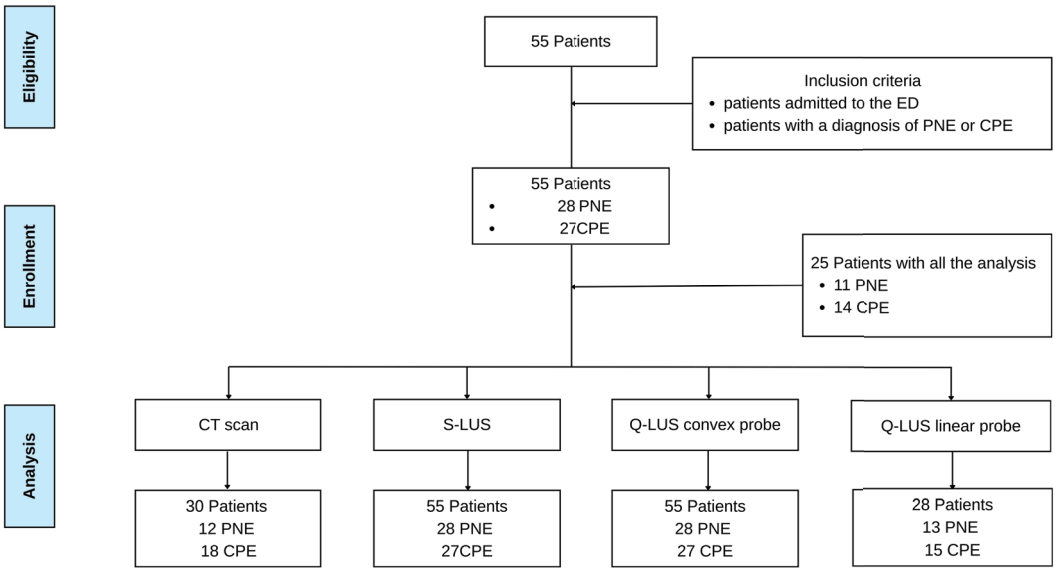

Figure 1. Consort flow diagram of the progress through patients that were enrolled, screened by imaging modality, and their data analyzed.

to a numerical semi-quantitative scoring-system that describes the severity of the pathological condition. However, this methodology represents a subjective operator-dependent practice, which inexorably leads to poorly reproducible outcomes. Moreover, S-LUS falls short in discriminating PNE from CPE, particularly because VA appear in both and look similar (7).

Ongoing research into the diagnostic capabilities of LUS has led to quantitative LUS (Q-LUS) (8-12). While for S-LUS the imaging patterns are analyzed qualitatively to assign a semi-quantitative score, with Q-LUS the patterns are characterized by objective measurements that correlate with the pathophysiological status of the lung. Recent studies focused on the spectral characterization of VA and their correlation with lung diseases. Mento et al. (12, 13) extracted spectral parameters such as native frequency, bandwidth, and $\max_f I_{TOT}$ to indirectly estimate the acoustic traps size, heterogeneity, and media content. The study was conducted on a cohort of 66 patients affected by CPE, PNE, or pulmonary fibrosis, to analyze Q-LUS parameters accuracy in differentiating diseases (12). While the results proved capable of differentiating between PNE and CPE, the study was performed without comparing findings to CT or S-LUS, nor did it investigate the importance of the total amount of areas scanned per patients. Thus, a knowledge gap remains: current lung imaging modalities lacks in objectivity, and spectral characteristics of VA are still underinvestigated. We hypothesize that Q-LUS could differentiate between cardiogenic edema (presenting acoustic traps filled with protein-poor fluid) and non-cardiogenic edema (acoustic traps filled with protein fluid and pus) (1, 4, 12). Starting from these hypotheses, we expect as the most indicative parameter the $\max_f I_{TOT}$, as it could characterize different absorption level in different edema compositions.

In this study, we evaluate and compare the diagnostic performance of four different lung imaging approaches. Precisely, CT scan and S-LUS, based on semi-quantitative scoring systems, and a Q-LUS approach based on the extraction of quantitative spectral features, with data acquired with a linear or a convex probe.

The data acquisition across the different imaging modalities and the statistical analysis are described in Sec. 2. The obtained results are shown in Sec. 3, whereas the conclusions are presented with the discussion in Sec. 4.

1. Methods

This is a prospective, single-center, observational diagnostic study conducted in the Emergency Department (ED) of Humanitas Gavazzeni Hospital, Bergamo, Italy, approved by the local Ethics Committee (approval number: 53/21 - 21/07/2021).

1.1 Dataset

Data were collected on adult patients presenting to ED with acute dyspnea or chest pain of suspected cardiopulmonary origin, between September 2021 and November 2022. From the study were excluded those patients presenting pregnancy, and refusal of consent. S-LUS and Q-LUS were performed within the first 6 hours after presentation in ED, while chest CT was performed only when clinically indicated, without a predefined order relative to S-LUS. Chest X-ray was performed based on clinical indication (28/28 in the PNE group and 22/27 in the CPE group). From admission, all patients underwent routine blood tests and arterial blood gas analysis. In total, 55 patients were enrolled, 28 affected by PNE (50.9%) and 27 with CPE (49.1%) (see Figure 1). The final diagnosis was derived by means of standard clinical examination by a panel of senior physicians integrating clinical, laboratory, and imaging data.

1.1.1 Chest CT scan

Chest CT was performed when deemed clinically necessary, using multidetector CT scanners available in the ED. Acquisition protocols followed institutional standards for thoracic imaging, with patients in supine position and scans covering the entire lung field (14) (see an example of CT scan image in Figure 2A). In total, CT scan data were acquired from 30 patients (54.5% with respect to the whole population), 18 affected by PNE (60.0%) and 12 by CPE (40.0%).

1.1.2 S-LUS

For S-LUS, the patients were examined in supine or semi-Fowler position. The examinations followed a standardized 12-zones acquisition protocol (posterior zones 2 and 5 excluded) (15-19). Each area was scanned with a Mindray M9 (Mindray Medical International Ltd., Shenzhen, Guangdong, China) scanner equipped with a Mindray convex probe C5-1s (frequency range: 2.3–5.7 MHz) and the cine-loops stored (12). To avoid potential bias, operators were blinded to other imaging and laboratory results at the time of acquisition. A total of 12 cine-loops for each patient were stored (see an example of S-LUS image in Figure 2B).

1.1.3 Q-LUS

Respect to S-LUS, Q-LUS presents a different acquisition setting and strategy. Indeed, with S-LUS, all the 12 areas were scanned with a clinical scanner, independently on the presence of VA. Differently, Q-LUS utilizes a programmable Ultrasound Advanced Research Platform (ULA-OP) to acquire radiofrequency (RF) data (20). Data were acquired dependently on the presence of VA as follows. Firstly, all the 12-zones were analyzed with the same settings utilized for S-LUS acquisitions; the zones presenting VA were successively scanned with ULA-OP connected to a convex probe (Esaote CA631) and the cine-loops were stored. Successively, the zones scanned with ULA-OP presenting VA were rescanned with the ULA-OP connected to a linear probe (Esaote LA533) and the cine-loops were stored. Here, the cine-loops stored are composed of 21 and 23 multifrequency images, centered at 3 (2, 3, and 4 MHz) and 4 (3, 4, 5 and 6 MHz) different center frequencies for convex and linear probe, respectively (20) (see an example of Q-LUS multifrequency images in Figure 2C and D). In total, all the 55 patients were acquired with the convex probe and 28 patients (15 CPE and 13 PNE) were acquired with the linear probe. The differences in the patient scanned with convex and linear probe are related to the original acquisition protocol, initially designed for a previous study (12).

1.2 Semi-quantitative scoring systems

1.2.1 Scoring system for CT scan

For each CT scan the following semi-quantitative scoring system was utilized to quantify parenchymal involvement across predefined lung regions (14). Each

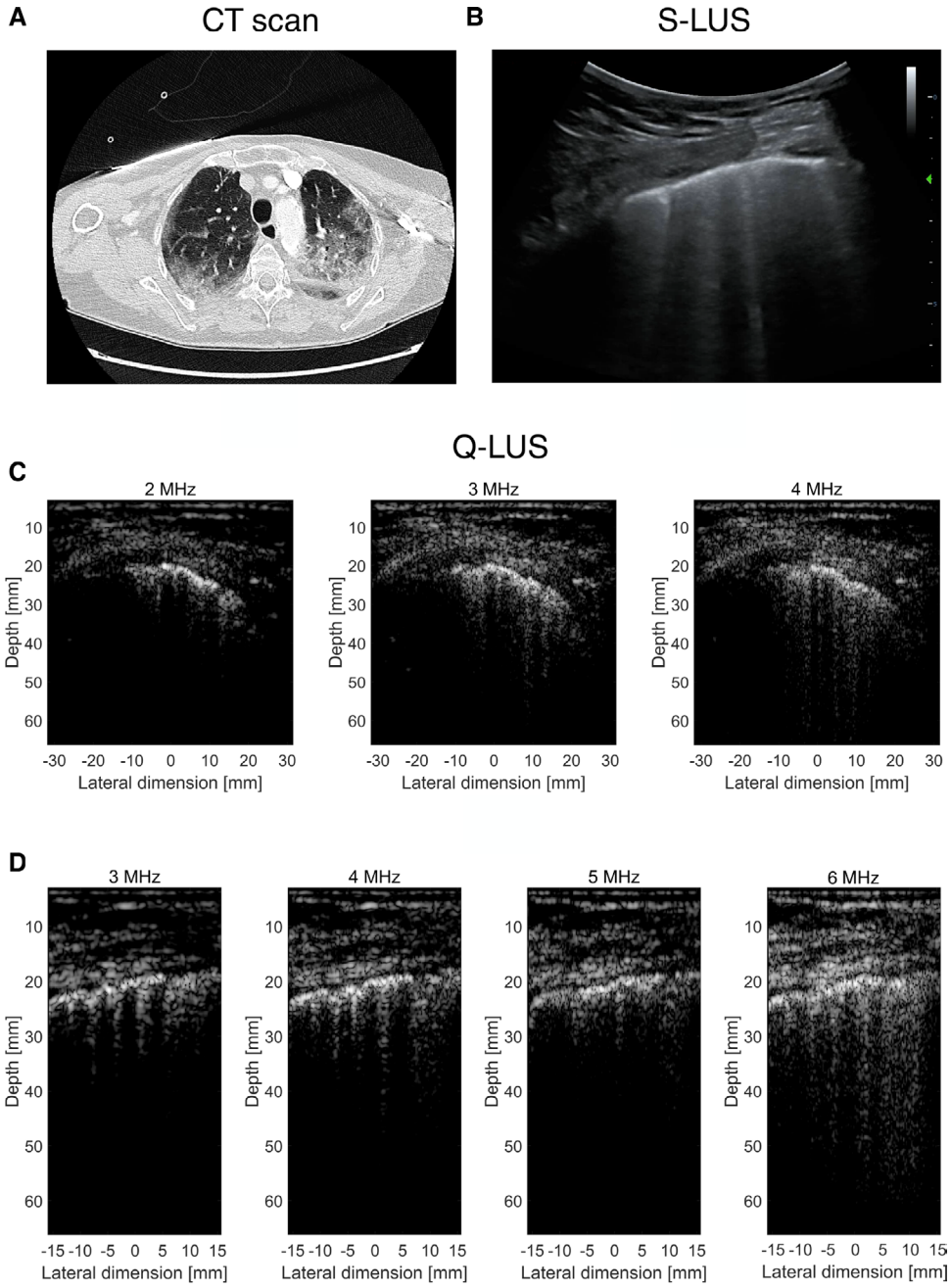

Figure 2. Images acquired by different imaging modalities. CT scan in a), S-LUS in b), Q-LUS acquired with the convex probe at 2, 3, and 4 MHz in c), and Q-LUS acquired with the linear probe at 3, 4, 5, and 6 MHz in d), both displayed at 35-dB dynamic range.

lung pair imaged through a CT scan was subdivided into 12 areas and the peripheral (cortical) lung zone of each area (defined as the outermost 2 cm of the parenchyma) was scored. The scores were assigned according to the predominant pattern observed: score 0, no evident pathological findings; score 1, initial pleural irregularities and/or early ground-glass changes; score 2, predominant ground-glass pattern and/or mixed areas with small consolidations; score 3, predominant consolidation pattern (not evaluable in cases of abundant

pleural effusion) (14). To avoid potential bias, the scores were assigned by a pneumologist and an emergency physician (with more than 10 years of experience in thoracic imaging), both blinded to clinical and ultrasound findings. Discrepancies between assessors were resolved by consensus between evaluating physicians. In total, 346 CT lung areas were scored (see Figure 3A) and used to characterize each patient in two ways: by the scores assigned to all the 12 areas or by the cumulative sum of the scores obtained by all the areas in a single patient.

1.2.2 Scoring system for S-LUS

S-LUS utilizes a semi-quantitative scoring system to provide a measure on the lung condition. A score from 0 to 3 is assigned to each of the twelve cine-loops stored. The score is assigned following the consensus guideline consisting on assigning score 0 to a video presenting a continuous pleural line with the possible presence of horizontal artifacts. Score 1 to a video presenting the first appearance of abnormalities, such as discontinuous pleural line and presence of VA. In score 2 the presence of small consolidations appears, reporting further loss of aeration on the lung. Score 3 presents a further deaerated lung presenting large consolidation and possibly VA (18, 21, 22). In total, 635 S-LUS areas were scored (see Figure 3B) and used to characterize each patient in two ways: by the scores assigned to all the 12 areas or by the cumulative sum of the scores obtained by all the areas in a single patient.

1.3 Q-LUS spectroscopy on vertical artifacts

1.3.1 Segmentation and quantification

VA quantification is performed with the following three-steps procedure (11, 12, 23, 24). In step 1, RF multifrequency images were post-processed by removing unwanted frequency components with a 12th order bandpass Butterworth filter (1.8-MHz bandwidth centered around each transmitted center frequency) (11, 12, 23, 24). Thereafter, the signal envelope was extracted by means of the Hilbert transform. Each multifrequency image was then normalized with respect to its maximum value (11, 12, 23, 24). At the end, the images were displayed in logarithmic scale with a 35-dB dynamic range (11, 12, 23). In step 2, VA were manually segmented by an operator. In this step it is important to clarify that a multifrequency image is composed of a subset of N images, where N is the number of selected center frequencies for each probe, 3 and 4 for linear and convex probe, respectively. Among each multifrequency image, the sub-image containing the most extended VA was selected to define the region of interest (ROI). For each VA, a ROI was manually defined with a width that fully contains the VA and a depth that starts 2-4 mm under the pleural-line and extends until the bottom of the image. The defined ROI is then applied to each subimage of a multifrequency image to extract N representations of the same VA at N different frequencies. Step 3, each segmented VA (one for each center frequency) was quantitatively measured by means of the $I_{TOT}(f)$ value (11, 12, 23, 25).

$$I_{TOT}(f) = 20\log_{10}\left(A_{PIX}\sum_{i,j}10^{\frac{ROI_{(i,j)}}{20}}\right)$$

A_{PIX} is the pixel area ($5.1018 \times 10^{-3}\ mm^2$ for the convex probe and $3.7877 \times 10^{-3}\ mm^2$ for the linear probe), i and j are the i-th row and j-th column of the pixel (with intensity greater than -35 dB) (12, 23, 26). $ROI_{(i,j)}$ is the pixel intensity in logarithmic scale.

1.3.2 Features extraction

Each VA is quantitatively described by three and four $I_{TOT}(f)$ quantities (one for each center frequency), for convex and linear probe, respectively. These quantities are utilized to extract three spectral features for each VA: native frequency, bandwidth, and $\max_f I_{TOT}$ (11, 12, 23). Native frequency is the frequency presenting $\max_f I_{TOT}$, bandwidth is the range of frequencies presenting I_{TOT} value above -6 dB from $\max_f I_{TOT}$, and $\max_f I_{TOT}$ is the maximum $I_{TOT}(f)$ of VA (11, 12, 23). The spectral features extracted are successively utilized to calculate minimum, 25% quartile (Q25), 50% quartile (Q50), 75% quartile (Q75),

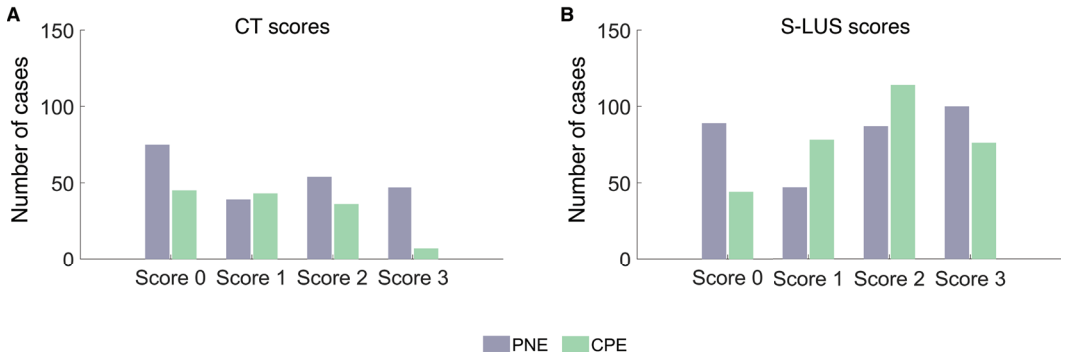

Figure 3. Histograms of a) CT scores and b) S-LUS scores for PNE (gray bars) and CPE (green bars).

and maximum values at the patient-level and at the area-level within a patient [e.g., at the patient-level, from a patient with 12 areas, we obtain for each patient a value for each feature; at the areas-level, we obtain for each patient 12 values for each feature]. This yielded two datasets of 15 statistical features, one at the patient level, and one at the area level, within a patient.

1.4 Statistical analysis

A comprehensive statistical analysis was performed by a biostatistician to compare the diagnostic performance of CT score, S-LUS score, and Q-LUS data with convex or linear probes. For CT and S-LUS, diagnostic accuracy was assessed using the conventional cumulative patient-level score. Complementarily, a Generalized Estimating Equations (GEE) at the lung-area level was applied to account for patient-level regional correlation. For Q-LUS, because each patient contributed multiple regions and each region contributed several artifacts, observations were modeled using GEE with clustering at patient ID to account the number of measures from a single patient. To identify the most informative quantitative features, we fitted a GEE logistic model for every possible non-empty combination of the 15 statistical features (2^{15}–1 total models); for each model, we computed the quasi-likelihood under the independence criterion (QIC), and selected, separately for convex and linear probes, the combination with the lowest QIC as the best-fitting model. For these best GEE models, we additionally fitted generalized linear mixed models including a random intercept for patient ID to quantify between-patient variability and derive the intraclass correlation coefficient (ICC). ICC was then computed to assess the degree of within-cluster dependence: values above 0.05 were considered indicative of non-independence and justified the use of a GEE framework. Multicollinearity among predictors was assessed using the multivariable Variance Inflation Factor (VIF), with values <4 regarded as acceptable. Only after confirming a suitable model family, non-negligible clustering, and acceptable collinearity, we proceeded with the GEE analysis. Odds ratios, 95% confidence intervals, and p-values were extracted and displayed through forest plots to compare feature relevance across probes. Diagnostic discrimination was evaluated using receiver-operating characteristic (ROC) curves and area under the curve (AUC), computed both at the patient and at the areas level for CT, S-LUS, and Q-LUS. We employed a repeated stratified nested k-fold cross-validation structured as follow. At the patient-level, an outer loop equal to 4 and an inner fold equal to 3 were utilized; at the area-level, an outer loop equal to 6 and an inner fold equal to 5 were utilized. Different fold dimensions between patient and area level were selected due to different data points available. Indeed, at the patient level we have 28 PNE and 27 CPE, for a total of 55 data points: at the area level this number increase, as each patient has multiple scanned areas. In the inner fold, GEE hyperparameters are optimized and then tested in the outer fold, where held-out predictions on patients never seen during training are produced. This procedure is repeated 10 times with different random fold assignments to mitigate potential

Table 1. List of the clinical analysis performed to find the diagnosis, divided into PNE (second column) and CPE (third column). The p-values of the distributions are reported on the third column.

Characteristics	PNE (N = 28[1])	CPE (N = 27[1])	p-value[2]
Sex			0.200
Male	21 (75.0%)	16 (59.3%)	
Female	7 (25.0%)	11 (40.7%)	
Average age [years]	80.5 (68.0–88.0)	88.0 (80.0–91.0)	0.059
ICU/PACU hospitalization [days]	3 (10.7%)	1 (3.7%)	0.611
Positive nasopharyngeal COVID-19 PCR	6 (21.4%)	0 (0.0%)	0.023
History of smoking	5 (17.9%)	5 (18.5%)	1.000
PaO_2/FiO_2 ratio [mmHg]	269 (238–362)	259 (209–326)	0.109
Nasal prongs			0.322
No	6 (21%)	9 (33%)	
Yes	22 (79%)	18 (67%)	
Non-invasive ventilation			0.937
No	21 (75%)	20 (74%)	
Yes	7 (25%)	7 (26%)	
Lactates [mmol/L]	1.20 (1.00–2.15)	1.20 (0.90–2.10)	0.375
White blood cells [10^3/µL]	12.2 (9.2–17.3)	7.9 (6.4–10.2)	0.003
C-reactive protein [mg/L]	0.56 (0.31–13.50)	0.16 (0.06–0.90)	0.019
B-type natriuretic peptide [pg/mL]	189 (100–457)	838 (500–1907)	<0.001
LUS Global Score	19 (11–25)	21 (12–26)	0.600
Consolidation on X-ray	18 (69%)	5 (21%)	<0.001
Interstitial diseases on X-ray	9 (35%)	19 (79%)	<0.002

[1]n (%); Median (Q1–Q3); [2]Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test. For each dichotomous variable (e.g., Yes/No), the table reports the number of patients meeting the condition (Yes) and the corresponding percentage within each diagnostic group.
Abbreviations: PNE = Pneumonia; CPE = Cardiogenic Pulmonary Edema; ICU = Intensive Care Unit; PACU = Post Anesthesia Care Unit; PaO_2 = Partial pressure of arterial oxygen; FiO_2 = Fraction of inspired oxygen; mmHg = millimetres of mercury; mmol/L = millimoles per litre; mg/L = milligrams per litre; pg/mL = picograms per millilitre; µL = microlitre.

partitioning bias. Since each patient contributes multiple observations (one per lung area), all observations from the same patient are kept together in the same fold. This prevents data leakage across folds. Missing CT or lung-area data were reported but not imputed.

2. Results

2.1 Semi-quantitative and quantitative data

Following the consort graph (27) in Figure 1, 30 patients were acquired with CT scan, for a total of 346 areas scored (average of 11.5 areas scored per patient), as depicted also in Figure 3A. Here, some areas present missing scores as the radiologist did not report them. Clinical characteristics of the overall study population are summarized in Table 1. Baseline demographic and clinical severity variables were comparable between the PNE and CPE groups. As shown in Table 1, all patients with pneumonia exhibited clear signs of systemic inflammation, with significantly higher white blood cell counts and C-reactive protein levels. In contrast, patients with heart failure displayed markedly higher indices of cardiac overload, particularly significantly elevated B-type natriuretic peptide (838 pg/mL). Notably, age, sex, days of intensive care unit hospitalization, degree of oxygenation, lactates, and oxygen support requirements were comparable between the two groups, indicating a similar overall

clinical severity. Furthermore, Table 1 reports imaging scores of CT scan and S-LUS and the features of chest X-ray. Overall, apart from imaging findings specific to each condition (e.g. consolidations on the chest X-ray, clearly suggesting PNE, or interstitial diseases, suggesting CPE), the main clinical and ventilatory characteristics were largely similar between the two populations. For S-LUS, the whole cohort of 55 patients were acquired, for a total of 635 areas scored (average of 11.5 areas scored per patient), as depicted also in Figure 3B. Here, some areas present missing scores as some scans were not performed due to the patient condition.

Q-LUS data with the convex probe were also acquired from the whole population, obtaining a total of 2216 VA (average of 40.3 VA per patient), which were segmented and quantified. Q-LUS data with the linear probe were acquired from 28 patients (50.9% with respect to the whole population), 13 with PNE (46.4%) and 15 with CPE (53.6%). From these patients, a total of 857 VA were segmented and quantified (average of 30.6 VA per patient). Differences in number of data acquired with the convex and linear probe are due to the acquisition protocol adopted for this dataset, which was acquired for a previous study (12).

2.2 Statistical analysis

In Table 2 the significance of the 15 statistical features is depicted for both convex and linear probes. From the convex probe, all the statistical features extracted from $\max_f I_{TOT}$ are significant, with a particular emphasis for the max of $\max_f I_{TOT}$, which shows the lowest p-value (≈ 0.0007). For the linear probe, 7 statistical features are significant, and well distributed between native frequency (2), bandwidth (3), and $\max_f I_{TOT}$ (3).

In Figure 4, the results of the QIC-based goodness-of-fit test depicts how the best performing Q-LUS model depends on the probe. For the convex probe, the optimal model included two features: Q25 of the native frequency and the max of $\max_f I_{TOT}$. For the linear probe, the best model combined Q50 native frequency, Q50 of $\max_f I_{TOT}$, and max bandwidth. Both best models showed minimal collinearity, as indicated by very low VIF values (all VIF ≤1.02), confirming the absence of multicollinearity among selected predictors. To quantify patient-level clustering, the best models were reestimated using logistic mixed models. Both probe configurations displayed extremely high intraclass correlation coefficients (ICC convex = 0.999; ICC linear = 0.998), indicating that most variance in Q-LUS features was attribulele to differences between patients rather than within-patient regional variability.

Table 2. P-values obtained from the Welch's t-test. The significant values (≤ 0.05) are highlighted in green.

	Features	Convex	Linear
NativeFrequency	Min	0.1492	0.0217
	Q25	0.0539	0.0991
	Q50	0.2925	0.0492
	Q75	0.4846	0.6681
	Max	0.5941	0.2167
Bandwidth	Min	0.1021	0.7536
	Q25	0.0868	0.1028
	Q50	0.1702	0.0011
	Q75	0.3260	0.0075
	Max	0.1369	0.0031
$\max_f I_{TOT}$	Min	0.0219	0.2272
	Q25	0.0068	0.0353
	Q50	0.0044	0.0148
	Q75	0.0021	0.0291
	Max	0.0007	0.0526

In Figure 5 the ROC curves are represented. Both convex and linear probe Q-LUS models clearly outperformed CT scan and S-LUS. AUC scores at the patient level showed values of 0.72 and 0.74 for Q-LUS with convex and linear probe, respectively; CT scan and S-LUS shows AUC of 0.65 and 0.53, respectively (Figure 5A). To further assess performance at a more granular resolution, ROC curves were reconstructed using area-level GEE models that accounted for patient-level correlation across lung regions (Figure 5B). Under this framework, the discrimination performance

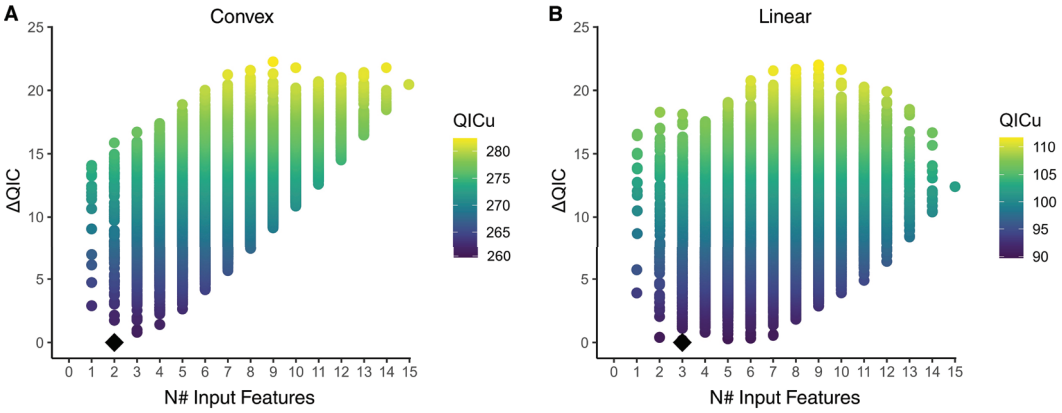

Figure 4. QIC of different GEE specifications using 15 features. Each dot represents a model; the color gradient indicates stability, with lighter shades denoting fewer stable models and darker shades indicating more stable ones. The x-axis shows the number of features included in each model, whereas the y-axis reports the ΔQIC, defined as the difference between the estimated model's QIC and the lowest QIC value obtained across all simulations. The black diamond marks the model we selected for both convex (a) and linear (b) probes.

of Q-LUS linear probe improved, with an AUC of 0.75, while with the convex probe the results show lower AUC values respect to the area level, 0.67; CT scan and S-LUS demonstrated poor discrimination performance with an AUC of 0.50 and 0.51, respectively.

3. Discussion and conclusion

Lung imaging serves as a support strategy in the diagnosis of lung diseases. Current imaging modalities lack diagnostic specificity and include gold-standard CT scan, which exposes patients to ionizing radiation, and S-LUS, which is based on the subjective interpretation of VA. These limitations are particularly relevant in conditions that share similar histopathological features, such as PNE and CPE. However, despite these similarities, their underlying pathophysiological mechanisms differ substantially and may be overlooked by CT and S-LUS standard imaging.

In this study, we looked at two histology presenting edema and, in line with expectations regarding the physical information provided by the features, the most important feature is the one who represent change in absorption. Indeed, as reported in Table 2, the statistical features of $\overset{max}{f} I_{TOT}$ are the ones who better distinguish PNE and CPE, which are similar from the point of view of their altered structure, but different in the content of edema. Therefore, these diseases are expected to differ in the way they absorb the ultrasound waves entering in the traps. Q-LUS demonstrated the ability to diagnose PNE and CPE through the quantification of VA patterns. In this paper, we compared the differential capabilities of CT scan, S-LUS, and Q-LUS convex and linear probes data acquired in-vivo from the same cohort of patients, affected by PNE or CPE. Results depict how Q-LUS outperforms CT scan and S-LUS in differentiating PNE and CPE in terms of AUC (see Figure 5). Q-LUS achieves the best performance when the linear probe is utilized, as the Welch's t-test, the QIC, and the ROC results depict, supporting the results obtained from previous studies (12, 13). Indeed, the linear probe offers a wider range of available frequencies that provide a more detailed characterization of the spectral features of a specific lung area (12, 13). With respect to the linear probe, the convex probe presents a limited range of frequencies, which limit the characterization of the lung areas (12, 13). The Welch's t-test support the results obtained from QIC, which depict as best features for discrimination of PNE and CPE, Q25 of the native frequency, and the max of $\overset{max}{f} I_{TOT}$ (the feature with the lowest p value ≈ 0.0007) for the convex probe. For the linear probe, the Q50 of the native frequency (p-value ≈ 0.0492) and $\overset{max}{f} I_{TOT}$ (p-value ≈ 0.0148) and max bandwidth (p-value ≈ 0.0031).

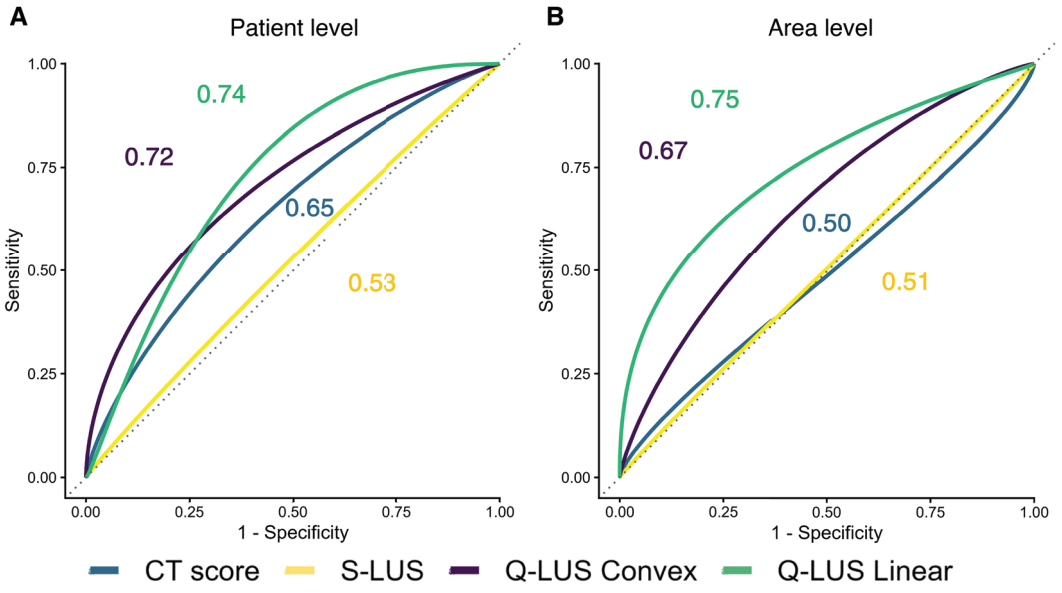

Figure 5. ROC curves comparing the GEE models with the lowest QIC and their corresponding feature sets for both the linear probe (green line) and the convex probe (purple line). Panel (a) displays ROC curves computed at the patient level, comparing these GEE models with the cumulative CT scan score (blue line) and the cumulative S-LUS score (yellow line). Panel (b) shows ROC curves comparing the same GEE models with regional scores within individual patients, both for the CT score (blue line) and the S-LUS score (yellow line). The numbers displayed correspond to the area under the curve (AUC) of each ROC.

CT and S-LUS imaging support for the diagnosis of lung diseases rely on subjective interpretation of imaging patterns. CT scan represents the gold standard, however, this imaging modality presents several limitations due to costs, radiations, and the need to transport patients to radiology. Chest CT was used in this study as the current clinical reference imaging modality rather than as an absolute gold standard for etiological discrimination between PNE and CPE. While CT remains highly sensitive in detecting lung involvement and its extent, its specificity in differentiating cardiogenic from non-cardiogenic pulmonary edema is known to be limited due to overlapping radiological patterns. In this context, CT represents the best available comparator in routine clinical practice, against which emerging imaging approaches should be tested, rather than a definitive ground truth. Accordingly, the inferior discriminative performance of CT observed in this study should be interpreted as a reflection of its intrinsic limitations in this specific diagnostic task, rather than as a methodological shortcoming.

S-LUS represents the less efficient tool in the diagnosis of lung edema, due to the subjective interpretation of the imaging patterns, the lack in diagnostic specificity, and the similarities between S-LUS findings in PNE and CPE. Our results markedly differ from those reported in the literature. Two recent meta-analysis studies showed high diagnostic performance for LUS, with a ROC ≥0.9 for pneumonia and acute heart failure. However, these papers also underscore that S-LUS relies on subjective interpretation of VA, which limits its specificity in distinguishing PNE from CPE (28, 29). Furthermore, the visualization of LUS patterns is dependent on parameters such as the acquisition frequency, focal depth, Time-to-Gain Compensation (TGC), and the type of probe. These factors limit the clinical outcome.

Results obtained with CT scan score and S-LUS score at the area level present low AUC values (≤0.51) suggesting a random discrimination of the two diseases. Differences between patient and area level remains invariate for S-LUS as also demonstrated in a randomized crossover study in ARDS patients (30). The regional S-LUS score reflects aeration in a specific lung area (0–3 per region) and shows substantial agreement with CT scan classification, whereas the

global S-LUS score sums regional values across 12 lung regions (0–36), providing an overall assessment of lung aeration and correlating strongly with mean CT scan density. However, the statistical approach in that trial was limited to paired tests and repeated-measures ANOVA without the use of GEE, despite the nested and repeated nature of the data. Such an approach may limit the robustness and generalizability of the findings, as it does not fully account for the hierarchical structure of the measurements (i.e., patient and lung areas). Indeed, the examination of single lung areas could reduce the ability to accurately describe the overall pathophysiological status of the lung. Instead, they could describe a different stage of the same pathology, or the presence of patterns formed by other causes (e.g. mechanical trauma) (31). Contrary, the areas-level method outperforms the patient-level method.

Spectral Q-LUS parameters (i.e., native frequency, bandwidth, and $\max_f I_{TOT}$) originate from the spectral spectral analysis of VA and offer a window into microscopic changes occurring at the lung surface. Their interpretation stems from the acoustic trap hypothesis: when the peripheral air-spaces are altered, they can trap ultrasound waves, and the characteristics of these trapped waves reveal information about the underlying tissue. Considering that the native frequency reflects indirectly the size of the trap dimension, bandwidth shows their homogeneity, and $\max_f I_{TOT}$ describe the filling content of the traps. With these measures, our best-performing model relied on the median native frequency and $\max_f I_{TOT}$, together with the maximum bandwidth, since these parameters capture the very essence of how the diseased lung reshapes and behaves. Here, we primarily associated different attenuations of the ultrasound backscattered signal to edema composition inside the acoustic traps: i.e., protein-rich and protein-poor fluids. Indeed, we expect that a protein-rich edema has a higher attenuation coefficient compared to protein-poor edema, leading to less intense VA. However, edema composition is not the only attenuating factor; other pathological elements (e.g., cellular debris) contribute with the alteration of the ultrasound wave propagation in complex ways. The presence of cellular debris or other anatomical components may hinder the ultrasound wave by generating additional multiple reflection and absorption phenomena, influencing the total intensity and the bandwidth of VAs spectral features.

It is important to clarify that the manual segmentation of vertical artifacts represents a subjective practice, which may lead to potential operator biases. Although this introduces a margin of subjectivity, Q-LUS is more objective than other imaging tools that present subjectivity in the detection of imaging patters and in the assignment of semi-quantitative scores. Additionally, to minimize operator-dependent variability, vertical artifacts were segmented by a single operator and successively verified and confirmed by expert clinicians with more than 15 years of experience in lung ultrasound. In future studies, to fully leverage Q-LUS objectivity, automatic segmentation algorithms need to be utilized. These algorithms need to be tested and optimized on larger amount of data to avoid further biases, overfitting, or poor performances.

Q-LUS with the linear probe is the most successful model, but it also presents the most limited number of samples, weakening the overall analysis. In future studies, additional patients, presenting both CT, S-LUS, and Q-LUS imaging exams, could be analyzed to strengthen the results of the analysis. We want to clarify that in daily clinical practice, the standard diagnostic pathway to discriminate between PNE and CPE is merely clinical: adding symptoms to imaging and response to drug/ventilatory therapy. Indeed, the focus of this study is not to differentiate PNE and CPE, but to evaluate and compare with the state of the art a new technology that supports the clinicians in three ways. First, the execution of a real-time and portable imaging modality, especially in hemodynamically and ventilatory insatiable patients. Second, a tool that enables monitoring of the lung pathology during time, in ways not possible with ionizing radiation exposure. Third, in case of overlapping pathologies (i.e., for PNE and CPE), the capability to distinguish different pathophysiological mechanisms.

To conclude, Q-LUS moves LUS beyond its traditional role as a bedside substitute for complex imaging. By reflecting real changes in lung microstructure, it provides information that are closer to the underlying pathology than current radiation-based or subjective

methods. This approach opens the door to safer, more accessible, and more biologically faithful monitoring of lung disease over time.

Funding: This work was supported by Hub Life Science-Advanced Diagnosis (HLS-AD), PNRR PNC-E3-2022-23683266 PNC-HLS-DA, INNOVA – CUP: E63C22003780001, funded by the Italian Ministry of Health under the National Complementary Plan Innovative Health Ecosystem - Unique Investment Code: PNC-E.3.

Ethics Approval and Consent to Participate: Emergency Department (ED) of Humanitas Gavazzeni Hospital, Bergamo, Italy, approved by the local Ethics Committee (approval number: 53/21 - 21/07/2021).

Data Availability: N/A.

Author Contribution: Mattia Perpenti: Investigation, Methodology, Data Curation, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization; Eleonora Balzani: Investigation, Methodology, Data Curation, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization; Federico Mento: Supervision, Writing - Review & Editing; Claudia Marinaro: Data Curation; Giacomo Bellani: Writing - Review & Editing; Tiziano Perrone: Conceptualization, Investigation, Supervision, Methodology, Resources, Writing - Review & Editing; Libertario Demi: Conceptualization, Investigation, Supervision, Methodology, Resources, Writing - Review & Editing, Project administration, Funding Acquisition

Declaration on the Use of AI: None.

Competing Interests: The authors declare that they have no competing interests.

References

1. Matthay MA, Zemans RL, Zimmerman GA, Arabi YM, Beitler JR, Mercat A, et al. Acute respiratory distress syndrome. Nat Rev Dis Primers 2019;5:18 doi:10.1038/s41572-019-0069-0
2. Mazzon E, Cuzzocrea S. Role of TNF-alpha in lung tight junction alteration in mouse model of acute lung inflammation. Respir Res 2007;8:75 doi:10.1186/1465-9921-8-75
3. Brito R, Lucena-Silva N, Torres L, Luna C, Correia J, da Silva G. The balance between the serum levels of IL-6 and IL-10 cytokines discriminates mild and severe acute pneumonia. BMC Pulm Med 2016;16:170 doi:10.1186/s12890-016-0324-z
4. Ware LB, Fremont RD, Bastarache JA, Calfee CS, Matthay MA. Determining the etiology of pulmonary edema by the edema fluid-to-plasma protein ratio. Eur Respir J 2009;35:331-7 doi:10.1183/09031936.00098709
5. Suciadi L, Pranata R, Huang I, Lim MA, Yonas E, Henrina J, et al. Comparing lung CT in COVID-19 pneumonia and acute heart failure: an imaging conundrum. Cureus 2021;13:e15120 doi:10.7759/cureus.15120
6. Mento F, Demi L. Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: an in vitro study. J Acoust Soc Am 2021;150:4075-82 doi:10.1121/10.0007482
7. Al Deeb M, Barbic S, Featherstone R, Dankoff J, Barbic D. Point-of-care ultrasonography for the diagnosis of acute cardiogenic pulmonary edema in patients presenting with acute dyspnea: a systematic review and meta-analysis. Acad Emerg Med 2014;21:843-52 doi:10.1111/acem.12435
8. Demi L, Wolfram F, Klersy C, De Silvestri A, Ferretti VV, Muller M, et al. New international guidelines and consensus on the use of lung ultrasound. J Ultrasound Med 2022;42 doi:10.1002/jum.16088
9. Mento F, Khan U, Faita F, Smargiassi A, Inchingolo R, Perrone T et al. State of the art in lung ultrasound, shifting from qualitative to quantitative analyses. Ultrasound Med Biol 2022;48:2398-416 doi:10.1016/j.ultrasmedbio.2022.07.007
10. Mento F, Perini M, Malacarne C, Demi L. Ultrasound multifrequency strategy to estimate the lung surface roughness, in silico and in vitro results. Ultrasonics 2023;135:107143 doi:10.1016/j.ultras.2023.107143
11. Perpenti M, Mento F, Pierro G, Perrotta A, Perrone T, Smargiassi A, et al. Fully automated quantitative lung ultrasound spectroscopy for the differential diagnosis of lung diseases: the first multicenter in-vivo clinical study. Comput Biol Med 2026;200:111365 doi:10.1016/j.compbiomed.2025.111365
12. Mento F, Perpenti M, Barcellona G, Perrone T, Demi L. Lung ultrasound spectroscopy applied to the differential diagnosis of pulmonary diseases: an in vivo multicenter clinical study IEEE Trans Ultrason Ferroelectr Freq Control 2024;71:1217-32 doi:10.1109/TUFFC.2024.3454956
13. Mento F, Perpenti M, Barcellona G, Perrone T, Demi L. Quantitative lung ultrasound spectroscopy classification performance in differentiating CPE, pneumonia, and PF: a comparative classifiers' analysis. IEEE Trans Ultrason Ferroelectr Freq Control 2024:1-4 doi:10.1109/UFFC-JS60046.2024.10793564
14. Smargiassi A, Soldati G, Torri E, Mento F, Milardi D, Del Giacomo P et al. Lung ultrasound for COVID-19 patchy pneumonia. J Ultrasound Med 2021;40:521-8 doi:10.1002/jum.15428

15. Mento F, Perrone T, Macioce VN, Tursi F, Buonsenso D, Torri E. On the impact of different lung ultrasound imaging protocols in the evaluation of patients affected by coronavirus disease 2019: how many acquisition do we need? J Ultrasound Med 2021;40:2235-8 doi:10.1002/jum.15580
16. Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, et al. Proposal for international standardization of the use of lung ultrasound for COVID-19 patients: a simple, quantitative, reproducible method J Ultrasound Med 2020;39 doi:10.1002/jum.15285
17. Mento F, Perrone T, Fiengo A, Tursi F, Macioce V. N., Smargiassi A, et al. Limiting the areas inspected by lung ultrasound leads to an underestimation of COVID-19 patients' condition. Intensive Care Med 2021;47 doi:10.1007/s00134-021-06407-0
18. Mento F, Perrone T, Fiengo A, Smargiassi A, Inchingolo R, Soldati G, et al. Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: a multicenter study. J Acoust Soc Am 2021;149:3626-34 doi:10.1121/10.0004855
19. Demi L, Mento F, Di Sabatino A, Fiengo A, Sabatini U, Macioce VN, et al. Lung ultrasound in COVID-19 and post-COVID-19 patients, an evidence-based approach. J Ultrasound Med 2022;41:2203-15 doi:10.1002/jum.15902
20. Tortoli P, Bassi L, Boni E, Dallai A, Guidi F, Ricci S. ULA-OP: an advanced open platform for ultrasound research. IEEE Trans Ultrason Ferroelectr Freq Control 2009 doi:10.1109/TUFFC.2009.1303
21. Jalilian H, Afrakhteh S, Mento F, Zannin E, Rigotti C, Cattaneo F, et al. Lung ultrasound video scoring using a novel motion-aware segmentation technique: toward automated neonatal LUS scoring. Comput Biol Med 2025;198:111244 doi:10.1016/j.compbiomed.2025.111244
22. Soldati G, Demi M, Smargiassi A, Inchingolo R, Demi L. The role of ultrasound lung artifacts in the diagnosis of respiratory diseases. Expert Rev Respir Med 2019;13:163-72 doi:10.1080/17476348.2019.1565997
23. Mento F, Soldati G, Prediletto R, Demi M, Demi L. Quantitative lung ultrasound spectroscopy applied to the diagnosis of pulmonary fibrosis: first clinical study. IEEE Trans Ultrason Ferroelectr Freq Control 2020:1-1 doi:10.1109/TUFFC.2020.3012289
24. Perpenti M, Mento F, Afrakhteh S, Barcellona G, Perrone T, Demi L. A novel empirical wavelet transform approach for classification of radiofrequency lung ultrasound signals applied to diagnosis of lung diseases. IEEE Trans Ultrason Ferroelectr Freq Control 2024:1-4 doi:10.1109/UFFC-JS60046.2024.10793884
25. Demi L, van Hoeve W, van Sloun R. J. G., Soldati G, Demi M. Determination of a potential quantitative measure of the state of the lung using lung ultrasound spectroscopy. Sci Rep 2017;7:12746 doi:10.1038/s41598-017-13078-9
26. Perpenti M, Mento F, Pierro G, Perrotta A, Smargiassi A, Inchingolo R, et al. Novel quantitative lung ultrasound spectroscopy approach for diseases classification. IEEE Trans Ultrason Ferroelectr Freq Control 2024:1-4 doi:10.1109/UFFC-JS60046.2024.10793543
27. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 explanation and elaboration: updated guideline for reporting randomised trials. BMJ 2025;389 doi:10.1136/bmj-2024-081124
28. Padrao E M H, Caldeira A B, Gardner T A, Miyawaki I A, Gomes C, Riceto L J, et al. Lung ultrasound findings and algorithms to detect pneumonia: a systematic review and diagnostic testing meta-analysis. Crit Care Med 2025;53 doi:10.1097/CCM.0000000000006818
29. Staub L J, Biscaro R R M, Kaszubowski E, Maurici R. Lung ultrasound for the emergency diagnosis of pneumonia, acute heart failure, and exacerbations of COPD/asthma in adults: a systematic review and meta-analysis. J Emerg Med 2018;56 doi:10.1016/j.jemermed.2018.09.009
30. Chiumello D, Mongodi S, Algieri I, Vergani G L, Orlando A, Via G, et al. Assessment of lung aeration and recruitment by CT scan and ultrasound in acute respiratory distress syndrome patients. Crit Care Med 2018;46:1 doi:10.1097/CCM.0000000000003340
31. Akor E A, Gao J, Guo J, Han B, Cruz A F, Herrmann J, et al. Structural and functional characteristics of healthy and injured porcine lungs during deflation: a quantitative CT imaging analysis. J Appl Physiol (1985) 2025;138:1615-27 doi:10.1152/japplphysiol.00443.2024