UNIVERSITÀ DEGLI STUDI
DI TRENTO

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
IECS International Doctoral School

Development of Lung Ultrasound Quantitative Approaches and Automatic Semi-Quantitative Strategies

In Silico, In Vitro, and Clinical Studies

Federico Mento

Advisor
Prof. Libertario Demi
University of Trento

November 2022

To my family

Abstract

Lung ultrasound (LUS) is a relatively novel application of ultrasound technology, which is increasingly expanding since the nineties. However, contrary to standard ultrasound imaging, which was primarily developed for imaging noninvasively the anatomy of internal body parts, LUS is mainly based on the visual interpretation of imaging artifacts. Among which, the so-called vertical artifacts are particularly important as they correlate with various pathologies. The main limitations associated with this type of pattern analysis remain its subjectivity and limited reproducibility. Moreover, the understanding and exploitation of the mechanisms behind the genesis of vertical artifacts are just in their beginnings.

In this context, the most diffused and utilized strategies in LUS analyses are the so-called semi-quantitative techniques, which are based on the visual interpretation of LUS patterns, where a score is assigned based on visual interpretation of LUS patterns, which correlate with the state of lung. However, these techniques are strongly operator dependent. To this end, the use of artificial intelligence (AI) to automatically score LUS data could be instrumental to reduce the subjectivity in the evaluation of LUS patterns. For this reason, as a first novel contribution, we proposed a technique to automatically classify LUS videos by means of an aggregation strategy based on a deep learning (DL) frame-based classification. However, given the strong subjectivity of the task, it is not reliable to expect levels of agreement between AI and human operators at video or frame level around 90-100%. Indeed, the use of AI algorithms could lead to more reproducible analyses but cannot completely avoid subjectivity, as AI training remains based on the subjective labeling performed by clinicians.

Another important aspect to be considered in semi-quantitative techniques is the proper definition and standardization of acquisition protocols. As an example, the number and spatial distribution of areas of the chest to be scanned are often defined arbitrarily and not following an evidence-based approach. For this reason, after having proposed (in a study of March 2020 I coauthored) a standardized imaging protocol specifically designed for the coronavirus disease 2019 (COVID-19) patients based on 14 scanning areas, we evaluated the impact of changing the scanning areas on the evaluation of COVID-19 (second novel contribution) and post-COVID-19 (third novel contribution) patients.

In addition, to properly define the imaging settings to be used in a LUS examination and, in parallel, to develop quantitative LUS techniques specifically designed for lung, the mechanisms behind the genesis of vertical artifacts (whose comprehension is just in its beginnings) should be deeply understood. To better understand the dependence of vertical artifacts on imaging parameters, we performed two experimental studies (fourth and fifth novel contributions), where we assessed the dependence of vertical artifacts' intensity on different imaging parameters. On one hand, the presented results showed how there exist different confounding factors (e.g., focal point and angle of incidence of ultrasound beam) that should be reduced when developing a LUS approach. On the other hand, the results showed how a frequency characterization of vertical artifacts could be exploited to develop a LUS quantitative approach, as these artifacts seemed to be associated with specific reso-

nance phenomena. Specifically, the acoustic traps' theory suggested that vertical artifacts originate from multiple reflections of ultrasound waves trapped within channels that can form between alveoli when lung tissue becomes pathological. By exploiting this concept, the frequency characterization of these artifacts could be used to indirectly estimate the size of acoustic channels (or traps).

To further evaluate the possibility to estimate these channels' size with a multi-frequency approach, we performed a numerical study with the k-wave MATLAB toolbox (sixth novel contribution). The main advantage of in silico studies consists of the possibility to control the disposition of alveoli, which can be located at precise distances between each other. Therefore, with this kind of studies, it is possible to look for a correlation between the vertical artifacts' intensity as a function of frequency and the alveolar disposition.

In the final novel contribution of this thesis, we performed a clinical study in humans showing the potentiality to exploit a quantitative multi-frequency approach to differentiate patients affected by pulmonary fibrosis (PF) from patients with other lung pathologies. Specifically, the frequency characterization of vertical artifacts along with their intensity was able to differentiate patients with PF with a specificity and sensitivity of 92%.

In conclusion, quantitative approaches represent the future of LUS, as they could provide a physical metric able to characterize the lung surface by applying an acquisition technique specifically designed for the lung. Nevertheless, to develop these techniques, the genesis of vertical artifacts needs to be more deeply investigated and understood by means of controlled in vitro and in silico studies. In the meantime, semi-quantitative approaches based on image analysis techniques should be exploited to estimate the state of lungs by detecting and recognizing specific LUS patterns that do signal different levels of aeration. However, to reduce the impact of confounding factors, the standardization of the imaging protocols and scoring systems is essential.

Acknowledgements

Three years have passed from the beginning of this experience and, even though several things happened during these years, time seems to be flown. Only four months after the beginning of this new adventure the outbreak of COVID-19 pandemic has forced many people at home, thus completely changing our habits.

This event coincided with moving house together with my girlfriend Chiara, who was able to bear me during all these two and a half years lived at home with me and, obviously, also the previous seven years spent together. I would like to thank you to having helped me during these years and to continue to help me. For your patience and your kindness that you show to all the people, thank you. I would also thank you to share with me this new huge experience to be parents of our son, Cesare, I love you.

I would like also to thank my mother Edda, my father Francesco, and my brother Andrea to always support me during these years at the university, thank you.

I would like to thank my girlfriend's family, which has always made me feel included, thank you.

I would like to thank my three historical friends Mauro, Satu and Son. Even though we did not meet too often during these three years, we have always been for each other, thanks guys.

I would like also to thank all my colleagues met during all the years spent at the university, both at the bachelor, the master and the PhD, thank you all.

I would like to thank all the members of ULTRa lab, who always made me feel part of team, thank you all.

Last but not least, a special thank to Libertario, who has been always here to teach me something new. We faced many stuff in the last years: crazy work at home during the pandemic, especially the first months, journals' special issues, editors, reviewers, discussions, and many publications. Thank you for your support as a supervisor and, especially, as a person.

A lot of stuff happened these years and we will see how many other things will happen the following ones.

So far, it was nice, see you tomorrow.

Federico

Contents

List of Tables v

List of Figures vii

List of Abbreviations xi

List of Symbols xiii

1 Preamble 1

2 Introduction and review of the LUS literature 5
2.1 Background . . . 5
2.2 Image Analysis . . . 8
2.2.1 Model based techniques . . . 8
2.2.2 Deep learning techniques . . . 10
2.3 Investigation of artifact genesis and quantitative studies . . . 18
2.3.1 *In vitro* and animal studies . . . 18
2.3.2 *In silico* studies . . . 21
2.3.3 Quantitative LUS imaging, human clinical studies . . . 22

3 Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study 25
3.1 Introduction . . . 25
3.2 Materials and Methods . . . 27
3.2.1 Data acquisition . . . 27
3.2.2 Data labeling . . . 27
3.2.3 Video-level score aggregating labeled frames . . . 28
3.2.4 Video-level score aggregating labeled and segmented frames . . . 29
3.3 Results . . . 32
3.3.1 Video-level score aggregating labeled frames . . . 32
3.3.2 Video-level score aggregating labeled and segmented frames . . . 34
3.4 Discussion and Conclusions . . . 37

4 On the Impact of Different Lung Ultrasound Imaging Protocols in the Evaluation of Patients Affected by Coronavirus Disease 2019, How Many

Acquisitions Are Needed? **41**
4.1 Introduction . . . 41
4.2 Materials and Methods . . . 42
4.3 Results and Conclusions . . . 43

5 Lung Ultrasound in COVID-19 and Post-COVID-19 Patients, an Evidence-Based Approach **45**
5.1 Introduction . . . 45
5.2 Materials and Methods . . . 47
5.2.1 Study Design and Population . . . 47
5.2.2 LUS Acquisition Protocol . . . 49
5.2.3 Impact of Different Scanning Areas on Exam's Evaluation . . . 50
5.2.4 Longitudinal Study . . . 52
5.2.5 Prognostic Evaluation . . . 53
5.3 Results . . . 53
5.3.1 Impact of Different Scanning Areas on Exam's Evaluation . . . 53
5.3.2 Longitudinal Study . . . 56
5.3.3 Prognostic Evaluation . . . 57
5.4 Discussion . . . 57
5.5 Conclusion . . . 59

6 On the influence of imaging parameters on lung ultrasound B-line artifacts, *in vitro* study **61**
6.1 Introduction . . . 61
6.2 Methods . . . 63
6.2.1 Lung-mimicking phantoms . . . 63
6.2.2 Data acquisition . . . 64
6.2.3 Computation of the total intensity parameter to quantify the strength of B-lines . . . 66
6.3 Results . . . 69
6.4 Conclusions and Discussion . . . 73

7 Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: An *in vitro* study **75**
7.1 Introduction . . . 75
7.2 Materials and Methods . . . 76
7.2.1 Materials and data acquisition . . . 76
7.2.2 Quantification of vertical artifacts' strength . . . 79
7.3 Results . . . 80
7.4 Discussion and Conclusions . . . 85

8 Numerical study on lung ultrasound B-line formation as a function of imaging frequency and alveolar geometries **89**
8.1 Introduction . . . 89
8.2 Methods . . . 91

8.2.1 Imaging parameters . . . 91
8.2.2 Computational domains . . . 92
8.2.3 B-lines' intensity metrics . . . 94
8.3 Results . . . 96
8.4 Discussion and Conclusions . . . 99

9 Quantitative Lung Ultrasound Spectroscopy Applied to the Diagnosis of Pulmonary Fibrosis: The First Clinical Study 103
9.1 Introduction . . . 103
9.2 Methods . . . 106
9.2.1 Data acquisition . . . 106
9.2.2 Quantification of B-Lines' Intensity . . . 108
9.2.3 Features and Statistical Analysis . . . 109
9.2.4 Classifiers . . . 111
9.3 Results . . . 112
9.3.1 B-Lines' Initial Evaluation . . . 112
9.3.2 Classifiers' Performance . . . 113
9.3.3 Features' Statistical Analysis . . . 113
9.4 Conclusion and Discussion . . . 114

10 Conclusion 117
10.1 Final Discussion . . . 117

List of Publications 121

Bibliography 127

List of Tables

2.1 Details on image analysis publications 18

6.1 The main parameters (number of transmitting elements, focus, and frequency) set in the different acquisitions 65

7.1 The parameters set for each scanning area 77

8.1 Imaging array parameters, being the bandwidth computed at –6 dB from the spectral peak. The array aperture can be computed as dx·(Pitch·N-Kerf), where dx is the grid size. 91
8.2 Domain grid sizes (dx) and time sampling intervals (dt) for each imaging frequency. 92
8.3 –12 dB lateral resolutions (LR) for each imaging frequency. 92
8.4 Original parameters (D, s), digitally approximated counterparts (D', s'), and average channel spacing ($\overline{s'}$) of four numerical domains 94

List of Figures

2.1 Examples of LUS images acquired with convex and linear probes 7

3.1 Examples of frames labeled as scores 0, 1, 2, and 3 28
3.2 The video-level agreement as a function of threshold (first algorithm) . . . 30
3.3 The video-level agreement and disagreement with the optimal threshold (first algorithm) 30
3.4 Boxplot showing the exam-level agreement as a function of the frame-level threshold (first algorithm) 31
3.5 Exam-level agreement when the optimal threshold is applied to the first algorithm 32
3.6 The exam-based cumulative scores for each LUS exam (first algorithm) . . 33
3.7 The video-level agreement as a function of threshold (second algorithm) . . 34
3.8 The video-level agreement and disagreement with the optimal combination of thresholds (second algorithm) 35
3.9 Exam-level agreement when the optimal combination of thresholds is applied to the second algorithm 35
3.10 The exam-based cumulative scores for each LUS exam (second algorithm) . 36
3.11 Video-level performance when the normalized distribution is analyzed . . . 37
3.12 Video-level performance when scores 0 and 1 are treated as a unique score (score 0/1) 38

4.1 The overall distribution of assigned scores divided per specific area, percentage of scores assigned for each area and patient, and the level of agreement for all of the patients' scanning 43
4.2 The overall distribution of score (four systems), level of agreement between systems, distribution of each score in the posterior areas, and level of agreement between the 3 modified versions of system 4 44

5.1 Sankey diagram illustrating the distribution of the dataset characteristics . 47
5.2 Typical LUS image associated with each level of the scoring system 48
5.3 Graphs referring to LUS exams performed on COVID-19 patients and post-COVID-19 patients (distribution of scores) 51
5.4 The overall distributions of scores, divided per specific area and per each subgroup 52
5.5 Graphs referring to LUS exams performed on COVID-19 patients and post-COVID-19 patients (distributions considering the four systems) 54

5.6 The values of Δ for each scanning area and for each patient 55
5.7 The exam-based sum of scores for each LUS exam 56

6.1 Microbubble microscope image, a phantom representation, and a top view picture of a lung-mimicking phantom 64
6.2 Picture of the acquisition system 66
6.3 The main elements in the reconstructed image 67
6.4 Schema of the procedure applied to the final (averaged) images to compute the total intensity of each main B-line 68
6.5 Multi-frequency images of phantom 2 (B-line IV) with the focus at 15 mm and 17.5 mm 69
6.6 Multi-frequency images of phantom 1 (B-line I) and phantom 2 (B-line III) 70
6.7 The normalized I_{TOT} of phantom 1 (B-line I) as a function of frequency . . 71
6.8 The normalized I_{TOT} of phantom 2 (B-line III) as a function of frequency and the -12 dB lateral resolution 72
6.9 Boxplot showing the I_{TOT} variations caused by frequency, focus, and number of transmitting elements 72

7.1 The utilized phantom, probe, positioning system, rotation system, metallic plate, water tank, and the three directions (x, y, and z) 77
7.2 The main elements visualizable in the reconstructed image 78
7.3 The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of center frequency 81
7.4 The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of angle of incidence 82
7.5 The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of bandwidth 83
7.6 The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of focal point 84
7.7 The normalized I_{TOT} of all the nine vertical artifacts, acquired from the eight scanning areas 85
7.8 Boxplot showing the I_{TOT} variations 86

8.1 Example of numerical domain, complete view (top) and enlargement (bottom) 93
8.2 Example of reconstructed image (top) and corresponding central line (bottom), indicated in red. 95
8.3 I_{tot}^N as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 20$ μm. 96
8.4 I_{tot}^N as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 100$ μm. 96
8.5 I_{tot}^N as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 198$ μm. 97
8.6 I_{tot}^N as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 263$ μm. 98

8.7 I_{tot}^N as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 395$ μm. 98
8.8 Reconstructed images at 1, 2, 3, 4, and 5 MHz, referring to the numerical domain made of alveoli having diameter $D = 300$ μm and spacing $s = 263$ μm. 99
8.9 I_{tot}^N as a function of muscle percentage below the lung surface p 100

9.1 Procedure applied to the ultrasound data to compute the total intensity of B-lines 107
9.2 Number of B-lines for each patient 108
9.3 Gaussian distributions and 2-D plot representing the patients 110
9.4 Multi-frequency images of a fibrotic patient and a nonfibrotic patient . . . 111
9.5 Histograms of I_{TOT} and 2-D plot (after application of a 1.6-dB threshold to the mean I_{TOT}) 112
9.6 Performance of the selected SVM and decision tree 114

List of Abbreviations

2-D	two-dimensional
3-D	three-dimensional
AI	artificial intelligence
ANOVA	analysis of variance
AOI	angle of incidence
AR	axial resolution
ARDS	acute respiratory distress syndrome
AUC	area under the curve
BW	bandwidth
CF	center frequency
CI	confidence interval
CNC	computer numerical control
CNN	convolutional neural network
COVID-19	coronavirus disease 2019
CPE	cardiogenic pulmonary edema
CT	computed tomography
DB	Daubechies
DL	deep learning
DNN	deep neural network
ELU	exponential linear unit
FEV1	forced expiratory volume
FP	focal point
FVC	forced vital capacity
GBDT	gradient boosting decision tree
GLCM	gray-level co-occurance matrix
grad-CAM	gradient-weighted class activation map
HRCT	high resolution computed tomography
ICC	intra-class correlation
ICLUS-DB	Italian COVID-19 LUS Database
ILD	interstitial lung disease
IoT	Internet of Things
IPF	idiopathic pulmonary fibrosis
LR	lateral resolution
LSTM	long short-term memory
LUS	lung ultrasound

LUSWE	LUS surface wave elastography
MD	medical doctor
MI	mechanical index
ML	machine learning
MSE	mean-square error
NN	neural network
PAS	peripheral airspace
PCA	principal component analysis
PF	pulmonary fibrosis
PFT	pulmonary function testing
POCUS	point-of-care ultrasound
PPP	peak positive pressure
PRF	pulse repetition frequency
PVC	polyvinyl chloride
RBE	recursive backward elimination
ReLU	rectified linear unit
RF	radio frequency
ROI	region of interest
RT-PCR	reverse transcription polymerase chain reaction
SAAB	subspace approximation with adjusted bias
SNR	signal-to-noise ratio
SOFT	simplified optical flow transform
SSC	systemic sclerosis
SSD	Single Shot Detection
STN	spatial transformer network
SVM	support vector machine
SYM	Symlet
TGC	time gain compensation
ULA-OP	ultrasound advanced open platform
WHO	World Health Organization

List of Symbols

A_{pix}	pixel area
c_{medium}	speed of sound in the *medium*
D	actual alveolar diameter
D'	digitally approximated alveolar diameter
DL_{ex}	deep learning algorithm exam-based score
DL_{SC}	deep learning algorithm video-level score
dt	sampling interval
dx	domain grid size
E_{agr}	exam-level agreement
f_0	center frequency
$I_{tot}(i)$	non-normalized total intensity of the ith image
I_{tot}^{max}	maximum value of total intensity in the whole simulation set
$I_{tot}^{N}(i)$	normalized total intensity of the ith image
I_{tot} or I_{TOT}	total intensity
K_{tot}	total number of spacing within an acoustic channel
MD_{ex}	medical doctor exam-based score
MD_{SC}	medical doctor video-level score
p	sub-pleural muscle percentage
ROI_B	vertical artifact's region of interest
ρ_{medium}	volumetric mass density of the *medium*
s	actual spacing
s'	digital approximated spacing
$\overline{s'}$	average alveolar spacing
TH_{OPT}^{segm}	frame-level segmentation optimal threshold
TH_{i}^{segm}	frame-level segmentation threshold
TH_{OPT}	frame-level labeling optimal threshold
TH_i	frame-level threshold
V_{agr}^{all}	overall percentage of agreement at video-level

Chapter 1

Preamble

In this first part, the main objectives of the thesis will be summarized to provide the reader with fundamental information about the research problem faced during the PhD.

Important potentials for the application of ultrasound to diagnose and monitor lung pathologies exist. However, there are limitations caused by the absence of standardized methodologies to acquire and analyze lung ultrasound (LUS) data. Moreover, the impact of several technical parameters (i.e., transmitted frequency and focal point) have not been considered. These limitations strongly affect the reproducibility of LUS.

The main objectives of the research presented in this thesis consist in providing solutions for the aforementioned problem, to improve both the LUS reproducibility and the quality of diagnostic value that can be extracted from LUS analyses. Therefore, designing quantitative LUS methodologies, adding specificity to LUS analysis, and improving reproducibility represent the main targets of this thesis.

These objectives can be achieved by

1. working with clinical data acquired with the currently available scanners (semi-quantitative application based on image analysis);
2. designing new LUS methodologies to insonify lungs and new processing algorithms to analyze the acquired data (quantitative application based on the analysis of raw radio frequency data).

In the first case, it is not possible to improve specificity as LUS patterns are not specific (i.e., the same patterns can be observed in different lung pathologies), but it is possible to improve LUS reproducibility. This can be done by

1.1 standardizing the acquisition procedure also to reduce confounding factors linked to the impact of imaging parameters (e.g., probes and frequencies);

1.2 adding to the semi-quantitative scoring systems (i.e., systems where scores are assigned based on visual interpretation of LUS patterns) the current knowledge of the phenomena observable during a LUS acquisition (e.g., not to count vertical artifacts but simply evaluate the presence or absence of specific patterns in LUS images);

1.3 utilizing automatic techniques to evaluate LUS data (baseline evaluation common to all clinicians).

In the second case, the aim is to redesign transmission and reception schemes as well as the processing of LUS data. Even though this is more complex to achieve than the first case, there exists potential of applying quantitative techniques and adding specificity to the analysis. This can be done by performing

2.1 *in silico* studies, which provide a highly controllable environment but simplified;

2.2 *in vitro* experimental studies, which provide a controllable environment and more similar to the clinical studies (as the same instrumentation of clinical studies is utilized, e.g., probes);

2.3 *in vivo* clinical studies.

On one hand, this thesis contributes to the development of semi-quantitative techniques as follows.

- After having proposed (in a study of March 2020 I coauthored) a standardized imaging protocol and scoring system specifically designed for COVID-19 patients (based on 14 scanning areas) [1], the impact of changing the number and location of scanning areas on the evaluation of COVID-19 (Chapters 4 and 5) and post-COVID-19 (Chapter 5) patients has been evaluated. This was done to better define the optimal trade-off between a time-efficient and accurate LUS examination procedure.
- After having proposed (in a study of May 2020 I coauthored) an automatic frame-based scoring system [2], an automatic aggregation technique has been developed (Chapter 3) and tested (Chapter 5). The aim of this technique was to pass from a frame-level classification to a video-level classification, which is the kind of evaluation needed by clinicians. The results showed how LUS data acquired with the currently-available scanners can be used to support diagnosis of lung pathologies. Specifically, the presented method can be effectively used for an automatic stratification of COVID-19 patients. However, some limitations linked to the reproducibility of results remain. Indeed, as demonstrated in a recent publication I coauthored [3], even if it is possible to design automatic techniques performing as a clinician, it is not possible to exceed a given level of agreement, as the application remain operator-dependent. Nevertheless, the variability of the interobserver agreement evaluated among clinicians is higher compared to the variability evaluated between clinicians and automatic algorithms [3]. This confirmed how both the reproducibility and reliability of the method improves when using automatic algorithms.

On the other hand, this thesis contributes to the development of quantitative techniques as follows.

- The main progress regards the understanding of vertical artifacts. Specifically, both from the *in vitro* (Chapters 6 and 7) and *in silico* (Chapter 8) studies, the mechanisms behind the genesis of vertical artifacts seem to be related to specific resonance phenomena, rather than reverberations.
- As follows from the previous point, if these phenomena are linked to a resonance, it is possible to indirectly estimate the spatial distribution of alveoli. Therefore, this

is functional to the improvement of specificity. This is shown in Chapter 9, where the potential of exploiting the frequency characterization of vertical artifacts for the differential diagnosis of pulmonary fibrosis has been proven in a clinical study.

To further summarize the content of this thesis, in Chapter 2, a review of the LUS literature will be provided, also briefly presenting the novel contributions that represent the main part of this PhD thesis. The remaining parts of the thesis are organized as follows. Chapters 3, 4, and 5 will present three contributions on LUS semi-quantitative approaches. While Chapter 3 will be mainly dedicated to the description of an automatic semi-quantitative approach, Chapters 4 and 5 will be mainly focused on the evaluation of a new LUS acquisition protocol and scoring system. Starting from Chapter 6, four contributions on novel quantitative approaches will be presented. Specifically, Chapters 6 and 7 will describe two *in vitro* experimental studies, whereas Chapters 8 and 9 will be focused on an *in silico* and a clinical study on humans, respectively.

Then, after the description of these seven contributions, the conclusions of this thesis will be drawn jointly with an overall discussion of the current LUS literature (Section 10.1).

Chapter 2

Introduction and review of the LUS literature

This Chapter[1] *aims at introducing the reader to lung ultrasound (LUS) by describing the main LUS studies, starting from the first ones, which date back the 60's. This introductory part will hence revise the main technical studies in LUS literature, also including the articles that I authored (novel contributions of this PhD thesis), allowing the reader to better understand the context within which these articles have been published. Then, in the following chapters (Chapters 3, 4, 5, 6, 7, 8, 9), these articles will be entirely reported.*

2.1 Background

The ability of ultrasound waves to penetrate media having similar acoustic impedances (e.g., soft tissue) makes them particularly suitable for medical imaging. Moreover, the presence of similar speeds of sound in human body is fundamental to reconstruct the anatomy with ultrasound imaging. Specifically, these two characteristics allow clinicians to anatomically investigate the internal parts of human body in real time without exposing patients to ionizing radiations. However, the anatomical investigation of aerated organs is not possible as the standard ultrasound imaging assumptions of similar acoustic impedances and quasi-homogeneous speed of sound in the volume of interest are unmet because of the presence of air. Indeed, the acoustic impedance of air significantly differs from that of soft-tissues, causing ultrasound waves to be almost completely reflected when encountering an acoustic interface formed by these two media. This is extremely relevant in lungs, as they consist of millions of air-filled alveoli.

The first studies on the use of ultrasound imaging for lung investigation were conducted by Dunn and Fry in 1961, in which they attempted to estimate ultrasound absorption and reflection in lung tissue [4]. Similar studies focused on the acoustical characterization of lung tissue were performed in the following years both *in vitro* [5–10] and *in vivo*

[1] This Chapter appears in:
[J0] F. Mento, U. Khan, F. Faita, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "State of The Art in Lung Ultrasound, Shifting From Qualitative To Quantitative Analyses," in *Ultrasound in Medicine and Biology*, in press, 2022.

[11]. However, these studies did not spread to clinical practice because of the difficulty in achieving reproducible estimations.

On the other hand, the first clinical applications of lung ultrasound (LUS) date back to 1967, when Joyner *et al.* reported the possible existence of characteristic ultrasound patterns in patients with pulmonary embolism and pleural effusion [12, 13]. Fifteen years later, Ziskin *et al.* observed the so-called *"comet tail artifact"* in a patient with a shotgun wound in the abdomen [14]. These artifacts were described as *"dense horizontal reverberation echoes likely caused by the strong reflection of objects having significantly different acoustic impedances with the background medium"* [14]. However, given their vertical extent, these artifacts are generally referred to as *vertical artifacts* or *comet tail artifacts*. The *"comet tail artifact"* was indeed detected by the authors also in canine liver in correspondence of lead pellets and by imaging glass and metallic rods in a water tank [14]. The clinical relevance of these artifacts was then highlighted in 1983, when Thickman *et al.* observed them in several tissue-gas interfaces (e.g., diagphram/aerated lung interface) [15]. A similar artifact, called *"ring-down"*, was reproduced *in vitro* by Avruch and Cooperberg, who observed these vertical artifacts when a bubble tetrahedron formed by two layers of soapy bubbles (diameter from 1 to 7 mm) was imaged by an ultrasound transducer [16]. However, only in the 1990s the clinical use of these vertical artifacts (first called *"comet tail"*) started to rapidly increase thanks to Lichtenstein *et al.*, who observed a correlation between this artifact and the presence of alveolar-interstitial syndrome [17]. We should also mention an important previous study that compared the sonographic appearance of pulmonary infarctions with their pathological reports [18]. However, contrary to Lichtenstein *et al.*, who compared *in vivo* the sonographic signs with computed tomography (CT) anatomical images of patients [17], this study was performed *ex vivo* [18].

Following the work of Lichtenstein *et al.* [17], the use of LUS techniques based on visual interpretations of imaging artifacts spread rapidly in the clinical world. For example, LUS vertical artifacts (called *"comets"* or *"B-lines"*) were exploited to assess extravascular lung water [19, 20] and differentiate acute cardiogenic pulmonary edema (CPE) from acute respiratory distress syndrome (ARDS) [21].

In addition to the presence of artifacts, ultrasound can be also used to detect subpleural consolidations, which consist of anatomical findings that could be imaged at the lung surface [22]. Specifically, consolidations are areas where lung tissue is substantially deaerated, thus making the lung significantly dense. Indeed, in these cases, lung is partially characterized by acoustical properties similar to those of soft tissues, and can thus be anatomically imaged.

In clinical practice, the common approach adopted by clinicians consists of visually inspecting LUS videos to detect the above-mentioned patterns. These qualitative techniques were reviewed and included in an international consensus on LUS, which is currently the only consensus publication on LUS [23].

These approaches form the basis of the so-called semi-quantitative techniques, which nowadays represent the most diffused and utilized strategies in LUS analyses. These techniques are indeed based on the visual interpretation of LUS patterns, where a score is assigned depending on the observed patterns, which correlate with the state of lung. As

an example, the lowest score of a scoring system is generally associated with a continuous pleural line (i.e., the image representation of the acoustic interface formed by intercostal tissues and air within lungs) with associated horizontal artifacts (also known as A-lines), which generally correlate with a healthy lung [1]. Horizontal artifacts represent equally spaced horizontal repetitions of the pleural line, and their genesis is linked to the presence of two strong reflectors (the probe and the pleural line), which causes ultrasound waves to bounce between these two interfaces [24]. As a healthy lung is characterized by a high level of aeration, with alveoli close to each other, the pleural line forms an acoustic interface having a reflection coefficient of about 1, thus generating horizontal artifacts in the image [24]. Figure 2.1 contains examples of horizontal and vertical artifacts, as well as consolidations.

As the widespread use of these approaches accelerated with the recent outbreak of the coronavirus disease 2019 (COVID-19) pandemic, a significant part of the literature on semi-quantitative LUS is related to COVID-19 applications [1, 25–41]. Nevertheless, these techniques are generally heterogeneous and are influenced by confounding factors, such as imaging frequency, focal depth, and utilized probes, which are generally not mentioned in the publications [42]. As a consequence, findings obtained using different approaches are extremely difficult to compare [42]. Another main problem is related to the strong operator dependence in the evaluation of LUS patterns and, thus, to the assigned score.

These limitations could be mitigated by adopting a standardized imaging protocol and scoring system, including technical aspects such as imaging parameters [1]. Moreover,

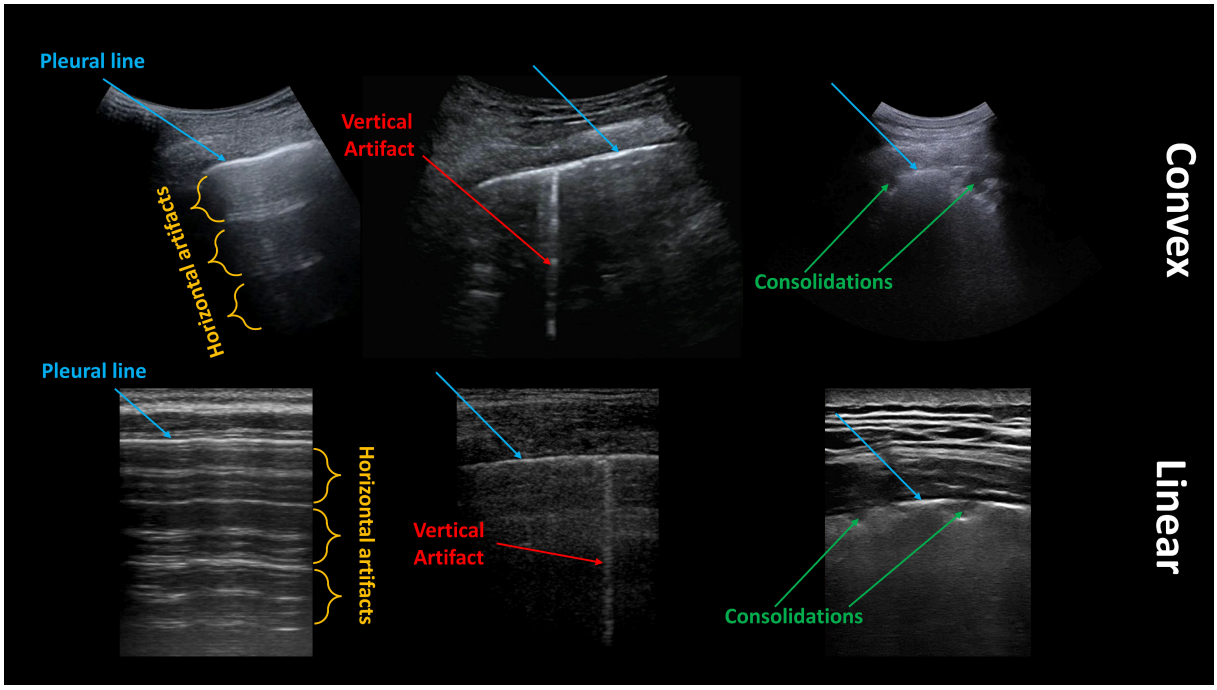

Figure 2.1: Examples of LUS images acquired with convex (top) and linear (bottom) probes are shown. Pleural lines, horizontal artifacts, vertical artifacts, and consolidations are indicated in blue, orange, red, and green, respectively.

artificial intelligence (AI) algorithms could be exploited to automatically classify LUS data based on scores, resulting in a more robust and reproducible semi-quantitative method [2, 43–48].

Because the necessity to develop acquisition strategies specifically designed for lung has been highlighted not only by clinicians, but also by researchers with technical backgrounds, the development of quantitative approaches could represent the future of LUS [49]. The evaluation of correlations between anatomical findings observed with standard imaging modalities (e.g., chest x-rays and CT), which are linked to histological observations, and LUS patterns [50–53] should represent an important starting point for developing ultrasound approaches specifically dedicated to lungs. This comparison could indeed allow researchers to study the link between the air spaces' (alveoli) resizing and spatial reorganization, as caused by different pathologies, and the quantifiable features of LUS artifacts [24, 54]. Then, these studies could be exploited to better comprehend LUS artifact genesis, allowing the development of quantitative LUS techniques aimed at estimating physical properties of lung surface [55–61].

However, the development of quantitative approaches requires the study and analysis of radiofrequency (RF) data; thus, these solutions are not yet available with clinical ultrasound imaging scanners [49].

In this introductory part of the thesis we will discuss the main technical publications on LUS, from semi-quantitative image analysis techniques (Section 2.2) to quantitative studies (*in vitro*, *in silico*, and *in vivo*; Section 2.3).

2.2 Image Analysis

In this section we will analyze the main contributions on LUS image analysis techniques. Specifically, model based techniques will be presented in subsection 2.2.1, and DL approaches in subsection 2.2.2.

2.2.1 Model based techniques

Contreras-Ojeda *et al.* proposed an approach to distinguish muscular tissues (above pleural line) from the artifactual structures below the pleural line in ultrasound images of healthy lung tissues of 13 pediatric patients [62]. Symlet (SYM) and Daubechies (DB) wavelet-based feature extraction, principal component analysis (PCA) and recursive backward elimination (RBE) were employed for feature selection, followed by K-nearest-neighbor based classification. This approach was able to achieve 97.5% of accuracy in discriminating muscular tissues from artifactual structures. Even though the results seem promising, the computational cost of the feature extraction phase could represent a relevant drawback for this application. This limitation could indeed strongly affect the advantage of real-time imaging of LUS.

Beyond that, various pathologies can be assessed by examining the pleural line, the subpleural regions and presence or absence of horizontal or vertical artifacts. To this end various techniques have been proposed.

In this regard, work has been done to automate the detection and localization of the

pleural line [43]. The authors proposed an unsupervised method based on the hidden Markov model and Viterbi algorithm. Furthermore, Support Vector Machines (SVM)-based classification was used to evaluate the characteristics of the pleural line, and the score value for each frame of a given LUS acquisition was assessed based on severity of the pathology. The proposed method was tested on a subset of the Italian COVID-19 LUS Database (ICLUS-DB), acquired from multiple clinical centers. Results revealed high image-based accuracy for both unsupervised detection of the pleural line and the classification performance. This study indicated the potential for real-time implementation, as the algorithms needed a total of 4 seconds (per image) to detect the pleural line and assign a score.

Anantrasirichai *et al.* described a method for restoration of lines in speckle images, followed by the automatic identification of vertical artifacts [63]. To that purpose, deblurring was performed in radon domain using a total variation blind deconvolution method. On the other hand, a local maxima technique in the Radon transform domain was used to identify the vertical artifacts. To evaluate line restoration and detection, 50 simulated images with varying image size were used. It was found that, for smaller regularization norms, noise from most of the lines was removed. Detection of vertical artifacts was evaluated using bedside data from 23 children. Although the approach showed promising results, it presents a high computational cost (45.75 seconds per image), making it infeasible for practical application. Moshavegh *et al.* presented a multistep study for automatic detection and visualization of vertical artifacts in LUS scans [64]. They identified the vertical artifacts as connected regions below the pleural line to the bottom end of the scan. Evaluation of 3,200 frames from healthy subjects and patients with pulmonary edema revealed that the average number of vertical artifacts was significantly higher in the patients than the healthy participants. Similarly, another method was presented by Karakus *et al.* [65]. This approach was formulated as a non-convex regularization problem involving a sparsity-enforcing Cauchy-based penalty function and the use of an inverse radon transform. The proposed method was validated in both radon and image domains, over examination of six male and three female patients. Results showed promising detection accuracy with improved identification performance of vertical artifacts when compared with the method of Anantrasirichai *et al.* [63]. However, this approach requires a processing time for a single frame of about 11–13 seconds, thus not allowing real-time imaging.

When discussing vertical artifacts, it is important to highlight how, given the strong dependence of vertical artifacts on imaging parameters [60, 66, 67], the count of vertical artifacts represents a qualitative application, which does not offer high reproducibility.

Moving from the analysis of vertical artifacts to examination of the pleural and subpleural space, a gray-level co-occurance matrix (GLCM) with second-order statistical textural-based digital analysis of pleural and subpleural space was proposed by Brusasco *et al.* [68]. They investigated the discriminating characteristics of a GLCM for differentiating ARDS from acute CPE. The analysis was performed on 47 participants (16 diagnosed with CPE and 8 with ARDS, and 23 declared as healthy). Results revealed statistical significance for 9 of 11 GLCM features on comparison of ARDS and CPE patient subgroups with the healthy participants. Similar statistical significance was reported

on comparison of ARDS with CPE. Conclusively, the proposed method showed potential in differentiating those with ARDS and CPE from healthy participants along with differentiating between patients with ARDS and CPE. Conclusively, it is again important to highlight how LUS analysis is highly subjective and that several confounding factors (e.g., imaging parameters and types of scanners) influence the appearance of the relevant imaging patterns. This leads to high variability among the generated images, affecting the reproducibility of these results. Moreover, the ground truth used to train AI algorithms remains based on subjective labeling of clinicians.

2.2.2 Deep learning techniques

Deep Learning (DL)-based analysis of LUS patterns can be applied to assess different pulmonary diseases and pathologies.

Kulhare *et al.* developed a convolutional neural network (CNN)-based algorithm to identify some of the lung features linked to pathological lung conditions. These features were defined as vertical artifacts, merged vertical artifacts, lack of lung sliding, consolidation and pleural effusion [69]. Ultrasound data from swine lung pathology models were captured for both normal and abnormal lungs in the form of 100 exams with 2,200 videos collected in total. The single-shot detection (SSD) framework was applicable to all LUS features, achieving at least 85% in sensitivity and specificity for all features. Lung features critical for diagnosing abnormal lung conditions were detected with greater than 85% accuracy.

To detect and localize the presence of vertical artifacts, Van Sloun and Demi presented a DL-based method for automatic detection and localization of vertical artifacts in an ultrasound scan [70]. A fully CNN was trained to perform this task on B-mode images of dedicated ultrasound phantoms *in vitro* and on patients *in vivo*. The *in vitro* study included 3,162 frames from 10 tissue-mimicking phantoms, while the *in vivo* study included 5,370 frames from 10 patients. Both showed high classification performance in localizing vertical artifacts. A gradient-weighted class activation map (grad-CAM) approach was used to guarantee a minimum level of explainability. Conclusively the method enabled detection and localization of vertical artifacts in real time.

Another method proposed by Kerdegari *et al.* was also aimed at automatically detecting and localizing vertical artifacts in LUS videos using DL networks trained with weak labels [71]. CNN combined with a long short-term memory (LSTM) network and a temporal attention mechanism was evaluated on LUS scans from 60 patients, totaling 298 examinations. These architectures varied in terms of convolutional networks (2-D or 3-D), presence and absence of temporal attention module along with the LSTM networks. The 2-D convolutional network with LSTM and attention module outperformed the other models, allowing the capture of features from both spatial and temporal dimensions. The model was able to achieve a classification F1-score of 0.81 with a localization accuracy of 67.1% within frames with vertical artifacts.

Baloescu *et al.* developed and tested a DL algorithm to assess the presence and absence of vertical artifacts in LUS [72]. A total of 400 consecutive thoracic ultrasound clips, each from a unique patient, were used. Each of the 400 clips was split into several 2,415 subclips, rated by two emergency physician point-of-care ultrasound (POCUS) experts. Rating

was based on a pre-determined ordinal scale from 0 (none) to 4 (severe), representing the number of vertical artifacts. When compared with expert interpretation for the presence or absence of vertical artifacts, the model for binary classification produced promising results with weighted kappa of 0.88 with 95% confidence interval (CI) between 0.79-0.97. Similar performance was also observed in the severity assessment using multiclass classification.

DL-based techniques have also been used to count the vertical artifacts present in the LUS scan. In this regard, Wang *et al.* proposed a study using a CNN to count vertical artifacts on a 4,864-image LUS dataset labeled by clinicians [73]. Furthermore, correlation between the automated count and the clinical parameters was examined. The clinical dataset was composed of 152 LUS videos corresponding to a total of 4,864 images. The available labels for each image ranged between 0 and 6, corresponding to the number of vertical artifacts in the image. The neural network (NN) agreed with the observer (*"true value"*) in 43.4% of the images, with an additional 40.8% in images with a deviation of 1. The intra-class correlation (ICC) for observer difference also revealed that the agreement between the human count and the output of the NN is high (ICC = 0.791).

Lung sliding is the respirophasic back-and-forth movement of the visceral and parietal pleural surface. Its presence indicates complete aeration of the lung at the site of probe placement on the chest wall, while its absence indicates the possibility of pneumothorax. DL-based techniques have been developed to automatically identify the presence or absence of lung sliding. In this regard, Mehanian *et al.* developed and compared several DL methods for identifying pneumothoraces in 3-second ultrasound videos collected with a handheld ultrasound system [74]. The first group of methods were based on CNNs paired with time-mapping pre-processing algorithms, including reconstructed M-mode and the proposed simplified optical flow transform (SOFT). The second class of algorithm used a DL architecture that combined CNN for processing spatial information (Inception V3) with a recurrent network (LSTM) for temporal analysis. To evaluate the methods, a total of four swine models were used, forming a total of 10 collection sessions per animal. As a result, a total of 130 positive videos with absence of lung sliding caused by pneumothorax and 122 negative videos with normal lung sliding were formed. The performance of the four methods in identifying the absence of lung sliding in swine pulmonary ultrasound videos was compared. All models learned informative representations of the data, all achieving area under the curve (AUC) values larger than 0.83 on unseen data.

Similarly, Jaščur *et al.* presented a novel DL-based automated M-mode classification method to detect the absence of lung sliding motion in LUS [75]. Automated M-mode classification leveraged semantic segmentation to select 2-D slices across the temporal dimension of the video recording. The dataset used to evaluate the study contained recordings of patients after thoracic surgery and were divided by physicians into two classes based on the presence and absence of lung sliding. They generated 17,338 frames. All presented models were pre-trained on the ImageNet database and then fine-tuned for a maximum of 15 epochs. The training and validation datasets followed the patient-wise split, and the ratio between them was kept to 1.5/1. Balanced accuracy ranged from 62% to 78%, with the best-performing model at 64 frames and the worst-performing model at 256 frames.

Pleural effusion refers to the buildup of excess fluid between the pleural layers outside the lungs. Tsai *et al.* aimed to develop an automated system for the interpretation of LUS of pleural effusion [76]. The standardized protocol followed involved scanning of six anatomical regions combined with a DL algorithm using a spatial transformer network (STN) providing the basis for automatic pathology classification on an image-based level. In this work, the DL model was trained using supervised and weakly supervised approaches, which used frame and video-based ground truth labels, respectively. In total, 623 ultrasound videos were acquired resulting in 99,209 2-D ultrasound images. To perform cross-validation, 10 folds of training and test sets were created, in which each patient appeared at least once in the test set. The video-based labeling approach reached 91.12% mean accuracy in the test set over the 10-folds, while the frame-based labeling approach reached 92.38%. In addition, a t-test on the accuracy of the two labeling approaches revealed no statistically significant difference in performance between the video-based and frame-based labeling approaches. This significantly reduced the input required from clinical experts to provide ground-truth labels.

Interstitial lung disease (ILD), appearing as fibrotic and stiffened lung parenchyma, may lead to symptoms, such as dyspnea, causing respiratory failure. Lung mass is not uniformly distributed in the lung, and it increases with the degree of fibrosis. Zhou and Zhang developed a method for analyzing lung mass density of superficial lung tissue using a deep neural network (DNN) and synthetic data of wave speed measurements with LUS surface wave elastography (LUSWE) [77]. Data were generated for fibrotic lung tissue, pulmonary congestion and edema, resulting in a total of 792,000 data measurements consisting of surface wave speed, excitation frequency, lung mass density, and viscoelasticity. Analyzing the convergence of different optimizers in terms of the validation loss over the epochs, the Adam optimizer had the highest validation accuracy, 0.992. When assessing the performance of the DNN (trained with synthetic data) with a sponge phantom, the predicted density from the DNN matched well the measured density of the sponge phantom, yielding an accuracy of 92%.

Zhou *et al.* proposed another study to develop a method for analyzing lung mass density of superficial lung tissue of patients with ILD and healthy participants using a DNN and LUSWE [78]. Surface wave speeds at three frequencies, predicted forced expiratory volume (FEV1% pre), ratio of forced expiratory volume to forced vital capacity (FEV1%=FVC%), age and weight of patients and healthy participants were used as features for training machine learning (ML) models. Random forest revealed that the contribution of age and weight were not as high as those of other features and, hence, these were not used in the training of the DNN model. As it was a retrospective study, 57 patients and 20 healthy participants underwent LUSWE. The evaluation of performance of the model was based on a train/validation/test (80/10/10) scheme. For the ReLU (rectified linear unit) and ELU (exponential linear unit), training and validation loss significantly decreased as the number of training epochs increased. Accuracies came out to be 84% and 89% for ReLU and ELU, respectively. Comparison of correlation coefficients with different activation functions in the DNN and ML for the testing dataset revealed that ELU in the DNN performed comparatively better.

As an extension to the previous study, Zhou *et al.* proposed *in vivo* prediction of

lung mass density for patients with ILD using different gradient boosting decision tree (GBDT) algorithms based on measurements from LUSWE and pulmonary function testing (PFT) [79]. The study used data similar to those used by Zhou *et al.* [78]. A fivefold cross-validation was conducted to assess the performance of different algorithms. Among the XGBoost, CatBoost, and LightGBM, mean-square errors (MSE) and correlation coefficients of the test dataset of three algorithms revealed that XGBoost obtained the best results.

DL models require high computational power and resources, making them unsuitable for deployment over lightweight devices such as mobiles and the Internet of Things (IoT). Furthermore, these models require efficient tuning of the hyperparameters. In the light of these concerns, Almeida *et al.* explored computer-aided assessment of pneumonia semiology based on light-weight NNs (MobileNets) [80]. Multitask learning was performed from online available COVID-19 datasets, for which semiology (overall abnormality, vertical artifacts, consolidations and pleural thickening) was annotated by two radiologists. The dataset consisted of a total of 12,718 images extracted from different LUS videos. A 75%/25%, train/test split at the image level was used for validation and testing. MobileNet outperformed the naive approach for all semiology indicators, with 95% accuracy for all semiology cases. Furthermore, in classification accuracy of MobileNet trained with labels provided by a senior radiologist, in comparison to labels independently provided by the junior radiologist, a high level of inconsistency was detected for mild conditions, with a mean accuracy of 77%.

On the other hand, Hou *et al.* proposed the use of an interpretable subspace approximation with adjusted bias (SAAB) multilayer network to screen the LUS images [81]. They demonstrated the advantage of using SAAB subspace network to design a low-complexity, low-cost, low-power-consumption solution for interpreting and visualizing features of LUS images to confirm the classifier recommendation. A dataset of 2,800 images was used for this study, consisting of 740 horizontal artifact images, 1,150 vertical artifact images and 910 consolidation images. Five hundred sixty images were used for testing, and 2,240 images were used for training. Greater than 96% accuracy over the testing data was obtained. In comparison to the CNN models, SaabNet needed to solve only 2,800 eigenvalues to yield an embedding vector of 1,183 elements and could be employed on any low cost simple board computers.

Erfanian Ebadi *et al.* proposed a method for fast and reliable interpretation of LUS images by use of DL, based on the Kinetics-I3D network [82]. The trained model could classify an entire LUS scan obtained at point of care, without requiring the use of preprocessing or a frame-by-frame analysis. The proposed video classifier was compared with ground-truth classification annotations provided by a set of expert radiologists and clinicians, which included horizontal artifacts, vertical artifacts, consolidation, and pleural effusion. A total of 1,530 videos were acquired corresponding to 287,549 frames. The models were trained and tested with fivefold cross-validation that creates training and testing sets with 80% (1,225 videos) and 20% (305 videos), respectively, for each fold. The model was able to produce a classification of an ultrasound video with 240 frames in 220 ms with accuracy score of 90%. The model learned to classify the severe disease cases (consolidation and/or pleural effusion) with a high F1-score.

Shifting from reducing computational complexity of the network to that of the data, Khan *et al.* proposed a method to analyze the impact of data compression on an automated scoring system [83]. The authors presented an automated scoring framework for reduced LUS data acquired from 20 COVID-19 patients, corresponding to 91,277 frames. LUS frames underwent spatial downsampling and reduced quantization levels by factors of 2, 3, 4 and 2, 4, 8, respectively. It was found that the prognostic agreement between expert LUS clinicians and the employed automatic algorithm ranged from 72.35% to 82.35% when reducing the data up to 32 times of its original size. This lays the foundation for efficient automated scoring in resource-constrained environments.

Roy *et al.* presented a novel fully annotated dataset of LUS image collected from several Italian hospitals, with labels indicating the degree of disease severity at the frame, video and pixel levels (segmentation masks) [2]. To evaluate the dataset they introduced several deep models that addressed relevant tasks for the automatic analysis of LUS images. In particular, they used a novel deep network, derived from STN, which simultaneously predicts the disease severity score associated with an input frame and provides localization of pathological artifacts in a weakly supervised manner. They also introduced a new method based on uninorms for effective frame score aggregation at a video level. The data were acquired from ICLUS-DB, which at the time included a total of 277 LUS videos from 35 patients (17 COVID-19-positive patients, 4 COVID-19-suspected patients, 14 healthy participants), corresponding to 58,924 frames. Data were split between test and train sets at the patient level; that is, the same patient was not included in both the train and test data sets. On the entire test set, frame level F1-score was 70.3%.

To further improve the DL performance, Frank *et al.* proposed a framework for training through integration of domain knowledge into DNNs by inputting anatomical features and LUS artifacts in the form of additional channels containing pleural and vertical artifact masks along with the raw LUS frames [47]. They used their framework to fine-tune a ResNet-18 model to classify each frame in its annotated severity score. The trained strategy and dataset were the same of Roy *et al.* [2]. The performance overcame the state of the art with an F1-score of 75.2%, highlighting the potential of domain knowledge integration to improve DL performance in LUS frame classification.

La Salvia *et al.* developed a system based on modern DL methodologies to automatically classify patients based on a seven-level scoring system [84]. For this purpose, they selected ResNet-18 and ResNet-50 architectures. Two thousand nine hundred and eight frames were carefully selected from a total of 5,400 videos (consisting of 60,000 frames) to train the models. The data were randomly split into training (75%), validation (15%) and test (10%) sets. By considering this split strategy and the amount of data (less than 3,000 frames), an accuracy above 96% was achieved.

Mento *et al.* reported on the level of agreement between DL models and LUS experts, when evaluating LUS videos [45]. As a result, they evaluated an empirical threshold approach to aggregate labeled frames to obtain a video-level score. The analyzed population consists of 82 COVID-19-positive patients corresponding to 314,879 frames. The overall video-level agreement reached its maximum value (51.61%) when a 1% frame-level threshold was applied. This means that a video was classified with the highest score appearing in at least 1% of frames in the video. A more relevant result was obtained when evaluating

the percentage of agreement between experts and DL in the stratification between patients at high risk of clinical worsening and patients at low risk. In particular, the approach achieved 85.96% agreement, thus highlighting the possibility of using DL approaches to automatically stratify COVID-19 patients. This aggregation strategy was later applied to a larger cohort of patients, comprising 100 COVID-19-positive patients and 120 post-COVID-19 patients (patients not positive at the time of LUS examination) [29]. Demi *et al.*'s study illustrated how the prognostic agreement was 80.45% for COVID-19 patients and 72.50% for post-COVID-19 patients [29]. The reduced performance on post-COVID-19 patients could be associated with the presence of LUS patterns not fully compatible with those obtained from healthy or acute patients [29]. Therefore, it is feasible that the AI models, which were trained on LUS data from COVID-19-positive patients, were not able to correctly recognize these patterns [29].

Xue *et al.* proposed a novel method for severity assessment of COVID-19 patients from LUS and clinical information [46]. Specifically, the authors stratified the task into three different steps. In the first step, they performed an LUS pattern segmentation at frame level by using a Visual Geometry Group (VGG) encoder. Then, in the second step, they classified LUS videos based on different features, including pattern segmentations obtained from the first step. Finally, they used the LUS score obtained in the second step and the patient's clinical information to assess the overall patient condition. The reported results showed how classification performance at the video level increased when pattern segmentations provided in the first step were exploited as additional input features. Furthermore, the performance at the patient level improved when video-level LUS scores (provided in second step) were included as additional features.

Chen *et al.* developed a technique to automatically classify LUS frames based on a scoring system [44]. A total of 45 patient were used to acquire 1,527 images assigned with scores and included in the study. Two different split strategies were tested. In the first, frames were randomly split between train and test sets, whereas, in the second, data were split at the patient level. In other words, the split was made to avoid having similar frames in the training and test datasets. The results showed how the performance obtained with the first strategy was higher than that with the second. Therefore, these results highlighted the importance of the splitting strategy when training an automatic algorithm.

Roshankhah *et al.* proposed an automatic segmentation method using a CNN to automatically classify LUS images based on a scoring system [48]. The study was evaluated by application of a randomly assigned and simple 90%/10% train set/test set split. Furthermore, the impact of splitting the training/test data was analyzed by repeating the process by performing the split between the train and test data at the patient level. The accuracy of the whole model at frame level was 95% when data were randomly split. In contrast, it was 68.7% when data were split at the patient level. These results, consistent with the study of Chen *et al.* [44], highlighted the importance of reliable splitting strategies when evaluating the performance of AI in classifying LUS data. Specifically, a random split at the frame level strongly affects the performance of AI algorithms, leading to overestimation of the capabilities of the automatic system.

The studies presented in this section are summarized in Table 2.1.

Publication	Study type	Amount of data	LUS patterns
[62]	clinical	22 images, NP videos, 13 patients	PL
[43]	clinical	3,315 images, 58 videos, 29 patients	PL
[63]	clinical	50 *in silico*, 100 clinical images, NP videos, 23 patients	PL, HA, VA
[64]	clinical	3,200 images, 64 videos, 8 patients	VA
[65]	clinical	100 images, NP videos, 9 patients	PL, VA
[68]	clinical	564 images, 564 videos, 47 patients	PL
[69]	animal	NP images, 2,200 videos, NP models	PL, VA, CON, LS
[70]	*in vitro*, clinical	3,162 *in vitro* and 5,770 clinical images, 10 *in vitro* and 27 clinical videos, 10 models and 10 patients	VA
[71]	clinical	NP images, NP videos, 60 patients	VA
[72]	clinical	28,980 images, 400 videos, 400 patients	VA
[73]	clinical	4,864 images, 152 videos, NP patients	VA
[74]	animal	NP images, 252 videos, 4 models	LS

[75]	clinical	17,338 images, 48 videos, 48 patients	LS
[76]	clinical	99,209 images, 623 videos, 70 patients	PE
[77]†	*in silico*, *in vitro*	792,000 parameters *in silico*, 165 parameters *in vitro*, NP models *in silico*, 1 model *in vitro*	none
[78]†	clinical	NP parameters, 77 patients	none
[79]†	clinical	NP parameters, 77 patients	none
[80]	clinical	12,718 images, 60 videos, NP patients	VA, PT, CON
[81]	clinical	2,800 images, NP videos, NP patients	HA, VA, CON
[82]	clinical	287,549 images, 1,530 videos, 300 patients	HA, VA, CON, PE
[83]*	clinical	91,277 images, 448 videos, 20 patients	HA, VA, CON
[2]	clinical	58,924 images, 277 videos, 35 patients	HA, VA, CON
[47]	clinical	58,924 images, 277 videos, 35 patients	HA, VA, CON
[84]	clinical	2,908 images, 5,400 videos, 450 patients	HA, VA, CON, PL
[45]*	clinical	314,879 images, 1,488 videos, 82 patients	HA, VA, CON

[29]*	clinical	772,780 images, 3,481 videos, 220 patients	HA, VA, CON
[46]	clinical	6,926 images, 1,791 videos, 313 patients	HA, VA, CON, PL
[44]	clinical	1,527 images, NP videos, 31 patients	PL, HA, VA
[48]	clinical	1,863 images, 203 videos, 32 patients	HA, VA, CON, PL

Table 2.1: Details on image analysis publications. The first column contains references to the publications. The second column indicates the study type (clinical, *in vitro*, *in silico*, or on animals). The third column reports the amount of data used in each study. If the amount of images, videos, or patients (models for *in vitro*, *in silico*, and animal studies) was not provided in the study, we use the acronym NP (not provided) in the table. The fourth column indicates which LUS patterns are investigated (PL = pleural line, HA = horizontal artifacts, VA = vertical artifacts, PE = pleural effusion, LS = lung sliding, PT = pleural thickening, CON = consolidations). The asterisk * indicates whether a clinical value exists in the study. For clinical value we here refer to the existence of a proven relation between the investigated LUS patterns and/or scores and the clinical state of the patient. The dagger † indicates a study concerning the assessment of parameters rather than images.

2.3 Investigation of artifact genesis and quantitative studies

In this section we will present the main LUS publications focused on investigation of vertical artifact genesis, and the main LUS quantitative studies. We will first discuss *in vitro* and animal studies (subsection 2.3.1), followed by *in silico* studies (subsection 2.3.2), and, finally, the quantitative LUS approaches tested in human clinical studies (subsection 2.3.3).

2.3.1 *In vitro* and animal studies

One of the first experimental *in vitro* studies was performed by Soldati *et al.* in 2011, where the appearance of LUS vertical artifacts was correlated with the density of wet, synthetic and partially aerated polyuretane sponges [85]. Specifically, the authors immersed 10 sponges (phantoms) in water and scanned them with a linear probe in five temporal phases during the drying process to visually observe the different LUS patterns appearing in the image. The scans revealed a transition from the first phase (completely wet sponge), characterized by a homogeneously echogenic field of view, also called *"white lung"*, to phases in which vertical artifacts appeared to be more spatially separated and, thus, the field progressively less echogenic. Therefore, for these models, LUS vertical

artifacts seem to be density correlated phenomena caused by modification of phantoms' porosity, which leads to different acoustic permeabilities [85]. In particular, a lower porosity (higher quantity of water filling air spaces) seems to induce greater acoustic permeability and greater acoustic interactions with the aerated superficial structures [85]. Similar results were obtained with *ex vivo* rabbit lungs, which were examined during mechanical inflation [86, 87]. Specifically, *ex vivo* deflated lungs were imaged with an ultrasound scanner at different levels of expansion, ranging from 87% (maximum lung expansion) to 40% (naturally collapsed lungs) of air content within lungs [86]. By decreasing the air content, hence increasing acoustic permeability, the observed LUS imaging patterns pass from a continuous pleural line with horizontal artifacts (associated to healthy lung) to the progressive presence of multiple vertical artifacts and, finally, *"white lung"* [86]. These results seem to confirm how acoustic lung permeability is a density-related phenomenon [86]. However, the authors believe that the investigation on vertical artifacts' genesis should be focused on the porosity in terms of shape and disposition, as altered density of peripheral lung can be simply seen as an epiphenomenon of altered peripheral airspace (PAS) geometry [87].

Following this hypothesis, Demi *et al.* designed an *in vitro* study to evaluate the possibility to characterize the lung structure (alveolar disposition and shape) by analyzing frequency spectra of vertical artifacts [55]. Lung-phantoms were produced by trapping a layer of monodisperse microbubbles in tissue-mimicking gel. Two different populations of phantoms were produced, one with 80-μm-diameter microbubbles and the other with 170-μm-diameter microbubbles [55]. These sizes were selected to mimic the alveolar size reduction (normal alveolar size equals 280 μm), which is typical of various pathologies [55]. The phantoms were scanned with a research platform by using a multifrequency approach in which images were sequentially generated using orthogonal subbands centered at different frequencies (3, 4, 5, ad 6 MHz with 1-MHz bandwidth) [55]. Vertical artifacts appeared with significantly stronger amplitude in specific portions of the frequency spectrum, highlighting the strong frequency dependence of these artifacts. This was also confirmed by an analysis on raw RF data [55]. These results suggest exploitation of the native frequency (i.e., the frequency at which the vertical artifact appears with stronger intensity) to characterize the state of the lung surface, thus opening to the possibility of developing a quantitative technique based on analysis of the artifact frequency spectrum [55]. To design such a quantitative technique, the dependence of vertical artifacts on different transmission parameters should be quantitatively evaluated. This was the aim of two recent quantitative studies, in which the dependence of these artifacts' intensity on different parameters was assessed in microbubble [60] and thorax [67] phantoms. The study on microbubble phantoms showed no correlation between the intensity of vertical artifacts and beam size, thus highlighting how changes in the lateral resolution do not affect the intensity of vertical artifacts [60]. Moreover, the center frequency was found to be the most impactful parameter in vertical artifact characterization, followed by focal point position and number of transmitting elements [60]. These results were confirmed by the second study, which also highlighted the importance of considering the impact of bandwidth and ultrasound beam angle of incidence when evaluating the artifact intensity [67]. These parameters indeed represent a significant source of variability in artifact evaluation

and, thus, should be carefully considered when implementing a quantitative approach [60, 67]. On one hand, center frequency could be used to characterize the state of lung surface, whereas, on the other hand, the other imaging parameters should be carefully set to prevent them from causing strong variations in the artifact intensity [60, 67].

A completely different quantitative approach was proposed by Mohanty *et al.*-a technique based on ultrasound multiple scattering that exploits the complex propagation of sound waves in the lung structure [56]. This approach is a near-field method based on assessment of the growth of a diffusive halo [56, 88]. Specifically, it aims at estimating diffusion constant and transport mean free path (L*) of lung parenchyma by using an acquisition approach in which the elements of an array were fired one by one [56]. Then, for each transmitted pulse, the received signals were collected on all the elements and these two parameters estimated [56]. This technique was tested *in vitro* with a sponge phantom with varying air volume fractions and both *in vivo* and *ex vivo* in rat lungs with induced pulmonary edema [56]. The results highlighted how a change of 10% of the air volume fraction corresponded to significant variation of L*. Indeed, when the quantity of fluid increases, the mean distance between scatterers increases, thus increasing L* [56]. In the following years, similar techniques were experimentally tested to differentiate rat lungs with pulmonary fibrosis (PF) from healthy lungs [89], detect pulmonary nodules (regions inside lungs containing no scatterers) [90] and differentiate rat lungs affected by PF (or edema) from healthy lungs of control rats [91].

In 2017 another group developed a quantitative approach based on LUSWE [57], which was then tested *in vitro* with a cellulose sponge filled with water in specific locations [92]. Specifically, water was injected at three locations of the phantom, and a shaker was placed in contact with its surface to generate harmonic vibrations at different frequencies (100, 150, and 200 Hz) [92]. An ultrasound probe was then placed at 0.5 cm from the shaker to detect and measure the waves propagating on the phantom surface [92]. The measurements were taken at six different stages: dry phantom and phantom injected with volumes of water from 3 to 15 mL (in 3-mL steps) [92]. The results indicated how surface wave speeds were generally higher when higher shaker frequencies were used, resulting in a behaviour similar to that observed in LUSWE measurements on human lung [92]. Nevertheless, by considering the six stages, the surface wave speeds appeared to be similar, highlighting the difficulty of assessing the quantity of water by means of the LUSWE technique [92].

As discussed in the preceding paragraphs, the design of lung phantoms represents an important step in the development of new LUS quantitative techniques and the investigation of vertical artifact genesis in a controlled environment. This was the main aim of a recent work in which different devices (phantoms) were fabricated by using agar cylinders (or disks), computer numerical control (CNC) milling machines and polyvinyl chloride (PVC) containers [93]. If imaged by an ultrasound probe, these phantoms were able to *in vitro* reproduce vertical artifacts, which changed their visual appearance when selected imaging parameters were differently set (e.g., pulse amplitude and center frequency) [93]. In a similar study, Kameda *et al.* produced simple experimental models able to generate vertical artifacts [66, 94]. Two different models were fabricated: the first consisted of different materials (i.e., a drop of ultrasound gel, a spindle-shaped juice sac of a mandarin

orange and a string shaped glucomannan gel) placed on a polypropylene sheet (simulating pleural line), and the second was made by placing glass beads and plates of different sizes on the sheet [66, 94]. The authors then imaged these phantoms by evaluating the visual effect of spatial compound imaging, focal point, center frequency, and probe type (convex vs linear) on the appearance of vertical artifacts [66, 94]. They qualitatively illustrated how the impact of these ultrasound machine settings on the vertical artifact morphology cannot be considered negligible [66, 94].

As mentioned above, the reproduction of LUS patterns in experimental phantoms has been proven to be an effective strategy to investigate their genesis. However, these artificial models could significantly differ from human lungs. Therefore, the use of a large animal model to reproduce LUS imaging patterns could represent an important tool to mimic the behaviour of human lungs. This was the scope of a recent study that assessed the usability of a pig model to reproduce LUS patterns of viral pneumonia [95]. Specifically, after anesthetization and intubation, six pigs were mechanically ventilated and a saline liquid was progressively instilled within their lungs using the one-lung flooding technique [95]. Then, the lungs were re-ventilated, imaged with an ultrasound scanner, and the acquired videos were classified based on a scoring system [95]. The LUS score increases (higher score corresponds to a worse state of lung) as the instilled saline fraction increases. Moreover, different LUS patterns associated to ARDS-related pneumonia (e.g., COVID-19 pneumonia) were observed [95]. Given the obtained results and the easy implementation of the model, it could represent an important instrument for LUS research, as it provides, in a more controlled environment, a model similar to humans.

2.3.2 *In silico* studies

Only recently researchers have started to perform studies on LUS *in silico*. The first study investigated the dependence of vertical artifacts on alveolar diameter and spacing (i.e., distance between alveoli) [96]. The numerical simulations were performed using the k-Wave MATLAB toolbox [97], which was used to replicate a simplified lung structure. Specifically, the first 2 centimeters of depth were composed of muscle tissue, whereas air inclusions (alveoli) were introduced beyond 2 centimeters [96]. These alveoli were periodically arranged to maintain the same distance between each other, both in the axial and lateral directions. The spacing and diameter of these air inclusions were then varied, and the volume of interest was imaged at different center frequencies (from 1 to 5 MHz with 1-MHz step size and 1-MHz pulse bandwidth) [96]. In particular, the spacing was set to 198, 263, and 395 μm, which correspond to half of the wavelength in muscle tissue at 4, 3, and 2 MHz, respectively. The quantitative results showed how the relation between artifact intensity and imaging frequency depends on the complex interaction between wavelength and alveolar geometries. By evaluation of the intensity when spacing was set to half of the wavelength, a possible correlation between artifact strength and the ratio between wavelength and spacing was found [96]. Moreover, as also proven by studies on rabbit lungs [86, 87], a higher density of lung (percentage of muscle in the air inclusion area) correlated with more intense vertical artifacts [96].

The second study aimed at reproducing *in silico* primary LUS patterns, such as horizontal and vertical artifacts [98]. The authors used a custom Fullwave numerical simulator

to model different acoustic traps, which were able to simulate different fluid portions in an affected lung region [98]. The area above the simulated lung consisted of simulated intercostal tissues, extracted by an optical human dataset [98]. The volume of interest was imaged with a simulated clinical phased array transducer. The results showed how these simulations were able to reproduce horizontal and vertical artifacts *in silico* [98]. Furthermore, consistent with other *ex vivo* [86, 87] and *in silico* [96] studies, the authors found a correlation between the density of lung (in terms of fluid portions in the alveoli) and vertical artifact appearance. Specifically, a 55% fluid portion is needed to detect vertical artifacts in these simulations [98].

2.3.3 Quantitative LUS imaging, human clinical studies

There have been few LUS clinical studies in humans aimed at testing or developing quantitative techniques. Indeed, only few researchers have recently moved in this direction. The first quantitative studies in humans were performed by Zhang *et al.*, who developed and applied LUSWE to estimate superficial lung tissue elastic properties [57]. Specifically, they aimed at differentiating patients with ILD and healthy participants by using LUSWE [57]. The implemented technique was the same at that applied by Zhou and Zhang for the *in vitro* study, in which a handheld shaker was used to generate a 0.1-s harmonic vibration on the phantom surface [92]. During the clinical study the measurements were taken on both lungs of 10 ILD patients and 10 healthy participants, who were examined through six intercostal spaces [57]. The lung surface wave speed was measured three times for each location, i.e., once for each frequency (100, 150, 200 Hz). The results showed significant differences on the surface wave speed between healthy participants and patients with ILD [57]. A similar study was conducted in 2019, in which the authors tried to use LUSWE for assessing ILD patients and systemic sclerosis (SSC) [99]. The technique was tested on 91 ILD patients (41 with SSC and 50 without SSC) and 30 healthy participants. As the previous study, the surface wave speeds of patient lungs were significantly higher than those of control participants. Nevertheless, no significant differences were reported between ILD patients with SSC and ILD patients without SSC [99]. The same group applied the LUSWE technique to perform two follow-up studies, aimed at assessing disease progression in patients with ILD [100] and pulmonary edema [61]. In both studies correlations between changes in lung surface speed and clinical assessments were found, highlighting the possibility of using LUSWE for quantitative assessment of ILD and edema progression [61, 100].

The other LUS quantitative approach that was tested in human clinical studies is based on LUS spectroscopy [58, 59], and was developed following the findings of a previous *in vitro* study [55]. Specifically, the theory underlying these studies consists of hypothesizing that vertical artifacts originate from the interaction between ultrasound waves and acoustic traps (or channels) formed at the lung surface when the lung becomes pathological [58]. The genesis of vertical artifacts seems indeed to be related to the formation of channels consisting of media that can be penetrated by ultrasound, such as blood, water and tissue [24, 54, 101]. When a sufficient quantity of energy enters these channels, it can be trapped and progressively irradiated toward the probe with a periodicity depending on the channel's size and the ultrasound pulse's frequency [58, 101].

Estimation of the sizes of these traps could be relevant in discriminating different pathologies, which are characterized by completely different alveolar dispositions and, thus, trap geometries. Specifically, the native frequency and bandwidth of vertical artifacts could carry information on, respectively, the size and geometry of the acoustic traps [58, 101]. By exploiting these concepts, Demi *et al.* presented an image formation protocol able to capture the frequency dependence of LUS vertical artifacts and visualize it in real time [58]. The final aim consisted of providing a quantitative evaluation of signals received from lung. Ten patients with various lung diseases (e.g., PF, pneumonia and adenocarcinoma) were enrolled, and RF data were acquired by means of a research platform. The acquisition strategy was based on a multifrequency approach in which four ultrasound images were formed with pulses having different center frequencies (3, 4, 5 and 6 MHz) [58]. The pulse repetition frequency (PRF) was fixed at 4 kHz to guarantee that the volume of interest did not significantly change when considering four images acquired with different center frequencies. Moreover, to evaluate more precisely the change in vertical artifacts along the frequency spectrum, a narrow bandwidth (1 MHz) was employed for each center frequency [58]. The results showed how the appearance of vertical artifacts significantly varied from patient to patient, and different native frequencies and bandwidths of these artifacts were estimated [58]. The multifrequency approach of Demi *et al.* [58] was later proposed by Mento *et al.* to differentiate PF from other lung pathologies [59]. Specifically, 26 patients were enrolled; half had idiopathic PF (IPF) and the other half were affected by different lung diseases (e.g., emphysema, pulmonary hypertension and asthma). The patients were examined by following the same procedure presented by Demi *et al.* [58], and RF data were acquired and analyzed [59]. One thousand twenty-nine vertical artifacts were detected and their main features (native frequency, bandwidth, and total intensity) were analyzed [59]. Results revealed how these three parameters could be exploited to discriminate patients with PF from patients with other lung diseases. In particular, when all three parameters were considered, an empirically defined binary classifier was able to achieve 92% specificity and sensitivity. Moreover, statistical analysis showed that native frequency and total intensity were significant in discriminating PF from other lung pathologies, wheareas the bandwidth was not [59]. The results obtained in statistical analysis were consistent with the results achieved when standard classifiers receiving these three features as input were employed to differentiate patients [59]. In conclusion, this study illustrated the potential to discriminate fibrotic patients by exploiting a quantitative approach based on the frequency and intensity analysis of vertical artifacts [59].

Chapter 3

Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study

This Chapter[2] *presents an automatic semi-quantitative approach for the classification of LUS videos. Specifically, starting from the frame-based classification of LUS images provided by pre-trained DL algorithms, we developed an empirical threshold approach to classify LUS videos. Indeed, even though these DL algorithms are able to automatically assign a score to each frame composing a video, clinicians need a video-level classification to evaluate the state of the patient. Therefore, we developed and tested an automatic threshold approach to pass from a frame-level classification to a video-level classification. In particular, the results on a population of 82 COVID-19 positive patients, examined according to a standardized imaging protocol and consisting of 1,488 videos (314,879 frames), showed how this approach, based on the output of DL models, can achieve high performance for the automatic scoring of LUS data.*

3.1 Introduction

As of today, February 15, 2021, a total of 108,484,802 coronavirus disease 2019 (COVID-19) confirmed cases and 2,394,323 deaths have been reported worldwide according to the World Health Organization WHO [102]. The Americas and Europe are the most hit WHO regions, with 48,401,821 and 36,573,613 confirmed cases and 1,143,432 and 812,370 deaths, respectively.

COVID-19 causes important complications to the respiratory system, and its symptomatology varies from patient to patient. Even though about 80% of patients show mild

[2]This Chapter appears in:
[J1] F. Mento, T. Perrone, A. Fiengo, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, "Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study," in *The Journal of Acoustical Society of America*, vol. 149, no. 5, pp. 3626–3634, May 2021.

symptoms, the other 20% shows serious complications that need hospital treatments [103]. In this context, both the monitoring of patients and a timely and correct diagnosis are needed [104]. The most currently used diagnostic tool is the reverse transcription polymerase chain reaction (RT-PCR) testing, which is performed by nasopharyngeal swab [105]. Nevertheless, this test shows a high false negative rate [106]. Another diagnostic tool is represented by chest computed tomography (CT), which has reported sensitivity between 61% and 99% and specificity between 25% and 33% [107]. Moreover, chest CT is not portable, uses ionizing radiations, and needs the patients to be moved within hospitals, thus increasing the contamination risk. In this context, lung ultrasound (LUS) has been proven to be an important ally for clinicians dealing with COVID-19 patients [32]. Thanks to its main characteristics (real-time imaging, portability, and wide availability), LUS nowadays represents an important imaging solution that can be applied for triage of symptomatic patients and patients' monitoring [32].

However, LUS is yet mainly based on the visual interpretation of imaging artifacts, i.e., the horizontal artifacts and the vertical artifacts, thus leading to qualitative and subjective diagnoses [24]. While the horizontal artifacts are characteristic of a healthy lung (which forms an acoustic interface named pleural line, which is highly reflective) [24], the vertical artifacts correlate with several pathologies affecting the lung parenchyma [21, 24, 50, 108]. The appearance of these vertical artifacts is likely caused by the replacement of the volume previously occupied by air with media acoustically similar to the intercostal tissues (e.g., tissue, water, or blood), which allows ultrasound waves to propagate beyond the pleural line [54]. Although there exist studies that aim at quantitatively evaluating the state of the lung parenchyma by means of ultrasound technology [55–60], these approaches are preliminary and not yet suitable for such an emergency context.

Nevertheless, even though a completely objective and quantitative analysis by means of LUS is not yet available, a proposal for the standardization of LUS in the COVID-19 context has recently emerged [1]. Specifically, a standardized imaging protocol and scoring system have been proposed, which enable the grading of the lung condition through a four level scoring system [1]. Different approaches, deep learning (DL) based [2] or signal processing based [43], have been recently developed to classify LUS frames according to this four level scoring system [1]. DL based algorithms have been developed on the basis of the above-introduced imaging protocol and scoring system [2], and are currently freely available to clinicians worldwide through a web-application [109]. These algorithms perform automatic scoring and semantic segmentation for each frame of a given LUS video, which normally contains hundreds of frames. However, clinicians in their practice evaluate LUS data, not at the frame level, but at the video-level. As an example, studies reporting the impact of different acquisition protocols can be mentioned [28, 30]. Moreover, a recent publication has demonstrated the prognostic value of the aforementioned imaging protocol and scoring system once evaluating the cumulative score, as obtained by adding the single scores on each of the 14 measurement points [38].

In this paper, expanding on the work on automatic frame-based evaluation, we thus evaluate the performance of the algorithms previously introduced when evaluating LUS data at the video and exam level (i.e., looking at the cumulative score on the 14 measurement points). Specifically, DL algorithms derived from Spatial Transformer Networks

(labeling algorithm) and from U-Nets and DeepLab v3+ (segmentation algorithm) [2] have been utilized to evaluate LUS data and the obtained results have been compared with the evaluation performed by expert clinicians (T.P. and A.S.). A total of 314 879 frames have been evaluated, corresponding to 1488 lung ultrasound videos acquired from 82 patients. All patients were COVID-19 positive as confirmed by RT-PCR swab test.

The paper is organized as follows. Both the data acquisition method and the studied population will be described in Section 3.2.1. Section 3.2.2 will be dedicated to the labeling procedure, whereas Secs. 3.2.3 and 3.2.4 will present the methods to obtain video-level scores from frame-level scores. Finally, the results will be presented in Section 3.3, and the conclusions, together with the discussion, will be summarized in Section 3.4.

3.2 Materials and Methods

3.2.1 Data acquisition

The analyzed population consists of 82 patients (43 male and 39 female, with ages ranging from 23 to 95 years, and average age equal to 61.1 years) with a diagnosis of COVID-19 positivity by a RT-PCR swab test. Of the 82 patients, 18 (8 male and 10 female, with ages ranging from 23 to 95 years, and average age equal to 52.1 years) were examined within the Fondazione Policlinico Universitario Agostino Gemelli (Rome, Italy), and 64 (35 male and 29 female, with ages ranging from 26 to 92 years, and average age equal to 63.7 years) within the Fondazione Policlinico San Matteo (Pavia, Italy). The patients were examined following the acquisition protocol presented by Soldati *et al.* [1] based on 14 scanning areas. As a subgroup of patients was examined multiple times (on different dates), a total of 114 LUS exams were performed (19 at Rome and 95 at Pavia), and 1,488 videos acquired (237 from Rome and 1,251 from Pavia), consisting of 314,879 frames (28,914 from Rome and 285,965 from Pavia). The data from Pavia have been all acquired using a convex probe with an Esaote MyLab Twice scanner, and an Esaote MyLab 50, setting an imaging depth from 8 to 12 cm (depending on the patient) and an imaging frequency from 5.0 to 6.6 MHz (depending on the scanner). The data from Rome have been all acquired using a convex probe with an Esaote MyLab 50, an Esaote MyLab Alpha, and a Philips IU22, setting an imaging depth from 8 to 12 cm (depending on the patient), and an imaging frequency from 3.5 to 6.6 MHz (depending on the scanner). This study was part of a registered protocol (NCT04322487) and received approval from the Ethical Committee of the Fondazione Policlinico Universitario Agostino Gemelli, Istituto di Ricovero e Cura a Carattere Scientifico (protocol 0015884/20 ID 3117), and the Fondazione Policlinico Universitario San Matteo (protocol 20200063198). All patients gave informed consent.

3.2.2 Data labeling

All the 1488 acquired videos were evaluated by expert clinicians (T.P. and A.S.), who assigned a score ranging from 0 to 3 to each video [1]. We referred to this evaluation as *"Medical Doctor (MD) score"*. Moreover, we fed the videos to the DL algorithms presented by Roy *et al.* [2] to obtain frame-level scores. As aforementioned in Section 3.1, we exploited two different DL algorithms: the first labeled each frame with a score

(we referred to these frames as *"labeled frames"*), and the second provided semantic segmentation, assigning one or more scores to each frame (we referred to these frames as *"segmented frames"*) [2]. The employed DL algorithms have been trained with the data set presented by Roy *et al.* [2], which is independent of the data set exploited in this article. It is important to highlight that data were split at patient level to avoid the presence of similar frames in train and test sets [2]. Figure 3.1 shows examples of frames labeled as scores 0, 1, 2, and 3.

3.2.3 Video-level score aggregating labeled frames

To obtain a DL video-level score from the aggregation of the DL frame-level scores, two strategies have been developed. While the first (presented in this section) is based only on the labeled frames, the second (presented in Section 3.2.4) combines the labeled frames with the segmented ones.

In particular, a threshold approach has been implemented and tested, which consists of assigning to a video the highest score assigned at least at a given percentage of frames (threshold) composing the video. Therefore, the target would consist in finding the optimal threshold (TH_{OPT}) able to maximize the overall percentage of agreement at video-level ($V^{all}_{agr}(TH_i)$) between the MD video-level scores ($MD_{SC}(p,a)$) and DL video-level scores ($DL_{SC}(TH_i,p,a)$), computed as

$$V^{all}_{agr}(TH_i) = \sum_{p=1}^{114}\sum_{a=1}^{14} \frac{100 \cdot V_{agr}(TH_i,p,a)}{1488}, \tag{3.1}$$

where the video-level agreement for a specific video (exam p and scanning area a) was computed as

$$V_{agr}(TH_i,p,a) = \begin{cases} 1, & \text{if } DIFF_{SC}(TH_i,p,a) = 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

with

$$DIFF_{SC}(TH_i,p,a) = MD_{SC}(p,a) - DL_{SC}(TH_i,p,a), \tag{3.3}$$

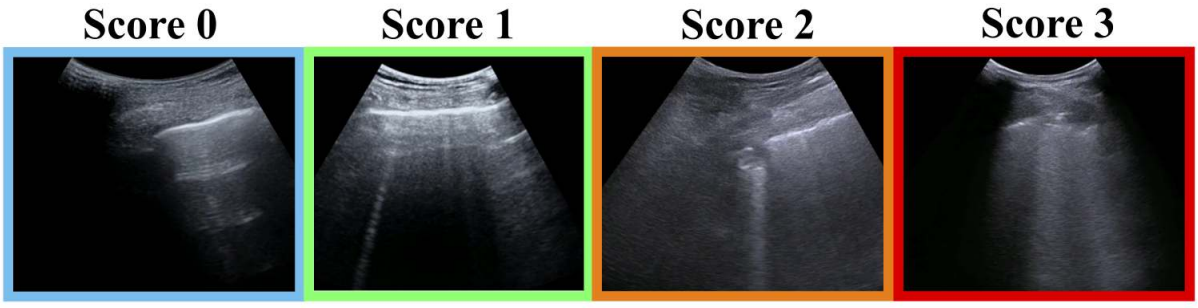

Figure 3.1: Examples of frames labeled as scores 0, 1, 2, and 3.

and TH_i being the frame-level threshold. It is noticeable how the total number of videos [denominator in Eq. 3.1] is equal to 1488 rather than $114 \times 14 = 1596$. In fact, during some exams it was not possible to scan all the 14 areas due to the patient's condition. To find TH_{OPT}, $V_{agr}^{all}(TH_i)$ was thus evaluated as a function of TH_i, which was varied from 0% to 100% with a 1% step size.

Then, TH_{OPT} only was applied to evaluate the performance of this aggregation technique in correctly scoring the LUS videos. It is clear how this simple approach implicitly allows the assessment of DL algorithms' performance [2] at video-level, starting from the frame-based classification. Therefore, by analyzing the percentage of videos having a DL video-level score different from the MD score, we implicitly evaluated what the most challenging scores are. For example, having a high percentage of videos labeled as score 0 by DL and as 3 by MD would mean that DL algorithms tend to misclassify score 3 as score 0.

Moreover, as previously mentioned in Section 3.1, the prognostic value of the scoring system [1] has been recently demonstrated [38]. Therefore, we statistically evaluated how the DL algorithms perform when looking at the cumulative score (exam-based score) ranging from 0 to 42. We indeed assessed the level of agreement between the MD exam based scores ($MD_{ex}(p)$) and the exam-based scores assigned by DL when applying the aforementioned aggregation technique ($DL_{ex}(TH_i, p)$), by computing the exam-level agreement ($E_{agr}(TH_i, p)$) for a specific exam p as

$$E_{agr}(TH_i, p) = MD_{ex}(p) - DL_{ex}(TH_i, p), \tag{3.4}$$

where $MD_{ex}(p)$ and $DL_{ex}(TH_i, p)$ are the exam-based scores of MD and DL for exam p, respectively, computed as

$$\begin{aligned} MD_{ex}(p) &= \sum_{a=1}^{14} MD_{SC}(p, a) \\ DL_{ex}(TH_i, p) &= \sum_{a=1}^{14} DL_{SC}(TH_i, p, a). \end{aligned} \tag{3.5}$$

Specifically, we focused on the DL performance at exam-level when TH_{OPT} is applied to each video.

3.2.4 Video-level score aggregating labeled and segmented frames

The aggregation strategy presented in this section is based on the combination of labeled and segmented frames. Specifically, a further step was added to the threshold approach based only on labeled frames (Section 3.2.3).

First, since the segmentation algorithm could provide multiple scores to each frame, we considered each segmented frame labeled only with the highest score assigned by the segmentation algorithm. We hence adopted a worst case strategy, as generally done in clinical practice. Then, we applied the same approach as Section 3.2.3 to find TH_{OPT} able to maximize the overall percentage of agreement at video-level. However, in this

case, the videos scored as 0 or 2 by DL when evaluating only the labeled frames were additionally processed by applying a threshold approach to the segmented frames. In

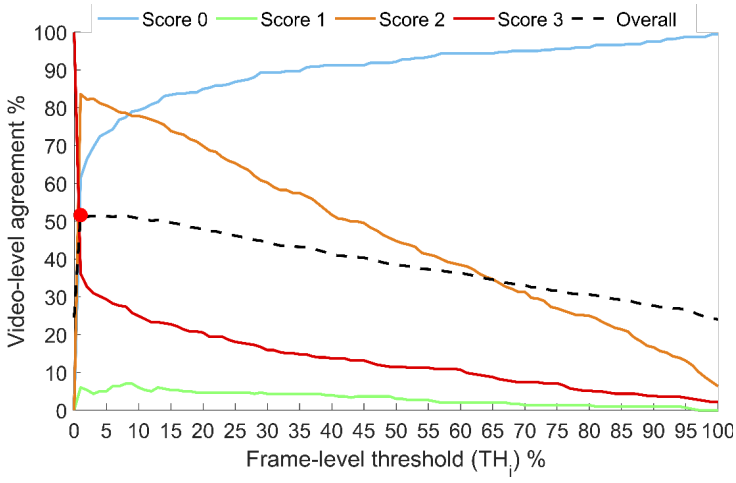

Figure 3.2: The video-level agreement as a function of threshold TH_i is shown (first algorithm, Section 3.2.3). The overall percentage of agreement at video-level ($V_{agr}^{all}(TH_i)$) is depicted with a dashed black line and the overall video-level agreement with the optimal threshold $TH_{OPT} = 1\%$ ($V_{agr}^{all}(TH_{OPT})$) is represented by a red point. The video-level agreement for scores 0, 1, 2, and 3 is indicated by light blue, light green, orange, and red lines, respectively.

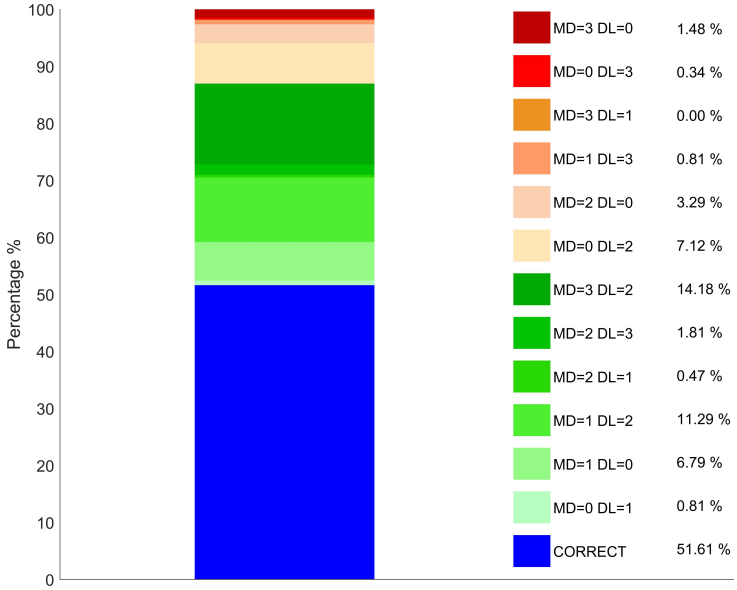

Figure 3.3: The video-level agreement and disagreement with the optimal threshold $TH_{OPT} = 1\%$ is shown (first algorithm, Section 3.2.3). The overall video-level agreement ($V_{agr}^{all}(TH_{OPT})$) is depicted in blue. The percentage of videos with disagreement between the MD scores and the DL scores is represented with different colors. A level of disagreement of 1, 2, and 3 points is represented by different tonalities of green, orange, and red, respectively.

particular, a frame-level segmentation threshold (TH_i^{segm}), varied from 0% to 100% with a 1% step size, was applied to these videos. We chose to test this approach in order to refine the aggregation technique presented in Section 3.2.3 for those scores (0 and 2) that were confused most often by the DL (see Section 3.3.1). It should be added that results showed us that only using the segmentation algorithm would have led to worse overall results in terms of agreement when compared to the aggregation technique presented in Section 3.2.3. Therefore, by using this second approach, we evaluated the overall percentage of agreement at video-level ($V_{agr}^{all}(TH_i, TH_i^{segm})$) as a function of TH_i and TH_i^{segm}), for a total number of 100 $\times$ 100 = 10 000 combinations of video-labeling. Specifically, we analyzed the optimal combinations of TH_i and TH_i^{segm} able to maximize $V_{agr}^{all}(TH_i, TH_i^{segm})$. Therefore, for each value of TH_i, there is a corresponding value of TH_i^{segm} able to maximize $V_{agr}^{all}(TH_i, TH_i^{segm})$ for that specific TH_i. We referred to this frame-level segmentation threshold as $TH_{OPT_i}^{segm}$. Once the optimal combination of thresholds ($TH_{OPT}, TH_{OPT}^{segm}$) was found, it was applied to each video to evaluate the performance at video-level and exam-level as done in Section 3.2.3.

Moreover, an analysis has been added by changing the number of analyzed videos. Specifically, the number of MD scores in the data set are not equally distributed, as the percentage of videos scored as 0, 1, 2, and 3 is about 21.44% (319 videos), 20.09% (299 videos), 33.94% (505 videos), and 24.53% (365 videos), respectively. Therefore, we normalized the distribution to have an equal number of videos scored by MD with

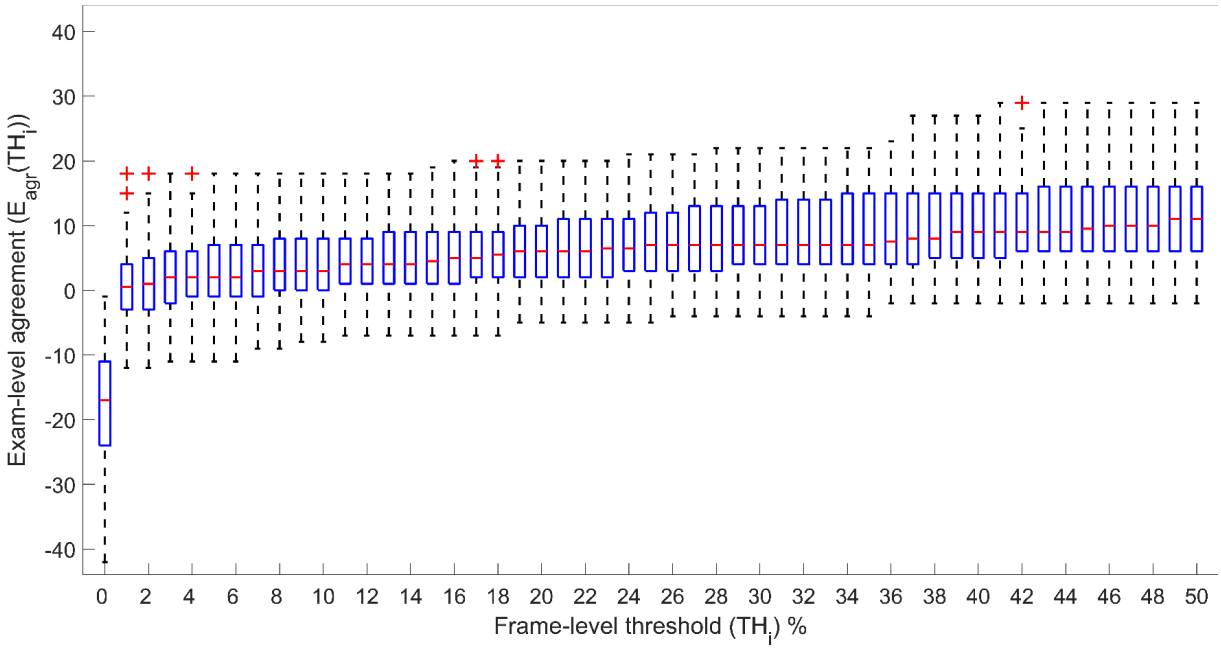

Figure 3.4: Boxplot showing the exam-level agreement ($E_{agr}(TH_i)$) as a function of the frame-level threshold TH_i, ranging from 0% to 50% with a 1% step size (first algorithm, Section 3.2.3). The median values are depicted with horizontal red lines, and the outliers are represented by red crosses. The lower and upper bounds of the boxes represent the 25th and 75th percentiles, respectively, whereas the maxima and the minima are represented by horizontal black lines.

each score (25% percentage of videos labeled by MD as scores 0, 1, 2, and 3). As a consequence, we selected all the videos scored as 1 by MD (i.e., the less frequent score), and then we randomly selected an equal number of videos (299) scored by MD as 0, 2, and 3, for a total number of 1196 analyzed videos. We hence assessed the performance of the DL aggregation technique at video-level by setting the previously-obtained optimal combination of thresholds $(TH_{OPT}, TH_{OPT}^{segm})$. This analysis allowed us to evaluate how the selected thresholds perform when considering a different number of videos, which have a normalized distribution of scores, and thus implicitly assessing the thresholds' reliability.

As final analysis, we evaluated the video-level performance of this aggregation technique (applying the optimal combination of thresholds $(TH_{OPT}, TH_{OPT}^{segm})$ when considering score 0 and score 1 as the same score, called *"score 0/1"*. In fact, score 1 represents a transition score (between a healthy lung and pathological lung with consolidations) that is characterized only by small alterations of the pleural line and is therefore challenging to be detected by DL algorithms. As a consequence, this analysis allowed us to assess the DL video-level performance in distinguishing a healthy or slightly altered lung (score 0/1) from a lung with advanced pathological conditions (score 2 and score 3).

3.3 Results

3.3.1 Video-level score aggregating labeled frames

Figure 3.2 shows the video-level agreement as a function of threshold TH_i. The overall video-level agreement $V_{agr}^{all}(TH_i)$ reaches its maximum value (51.61%) when a 1% frame-level threshold ($TH_i = TH_{OPT} = 1\%$) is applied, whereas its minimum value (24%) is

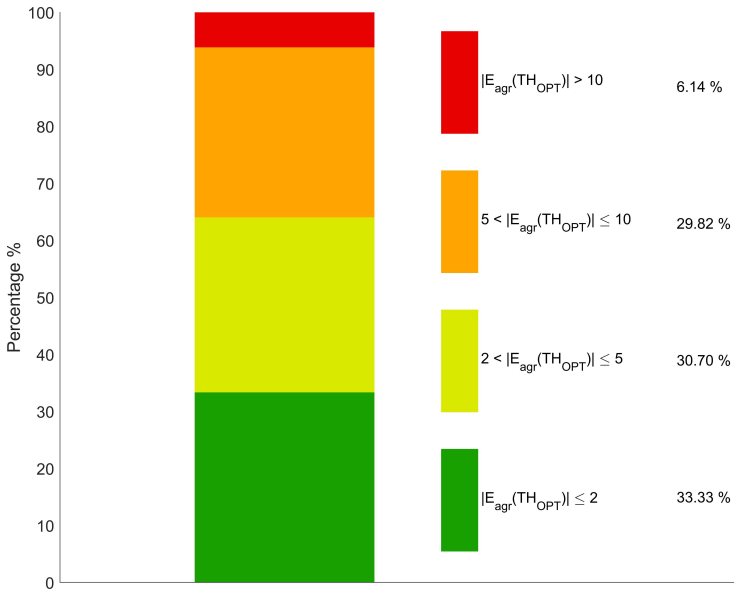

Figure 3.5: Exam-level agreement ($E_{agr}(TH_{OPT})$) when the optimal threshold is applied to the first algorithm (Section 3.2.3). The different colors represent the percentage of exams having a disagreement within a specific interval.

found when applying a 100% frame-level threshold. Moreover, $V_{agr}^{all}(TH_i)$ has a decreasing trend when evaluating values of TH_i from 1% to 100%. In contrast, when TH_i is set to 0%, the overall video-level agreement is low (24.53%), and corresponds to the percentage of videos scored as 3 by MD. In fact, when TH_i is set to 0%, all the videos are scored as 3 by DL, leading to a 100% video-level agreement for score 3, and a 0% video-level agreement for the other scores. Furthermore, the video-level agreement for score 0 has an increasing trend, whereas, in contrast, score 3 has a decreasing trend of video-level agreement. These behaviors were expected as, by increasing TH_i, a higher percentage of frames scored as 3 by DL are needed to score the video as 3. For this reason, also the video-level agreement of score 2 has a decreasing trend for values of TH_i from 1% to 100%. However, the score 2 video-level agreement is always higher (maximum value about 83.56% at $TH_{OPT} = 1\%$) than the score 3 video-level agreement in this range of TH_i (maximum value about 36.16% at $TH_{OPT} = 1\%$). The poorest performance is observed when evaluating the score 1 video-level agreement, which ranges between 0% and 7.02% (6.02% at $TH_{OPT} = 1\%$).

Figure 3.3 shows a perfect agreement $V_{agr}^{all}(TH_{OPT})$ for 51.61% of the videos, which increases to 86.96% when allowing a disagreement up to 1 point. However, for a significant number of videos, DL tends to misclassify score 3 as score 2 (14.18% of videos), score 1 as score 2 (11.29% of videos), and score 0 as score 2 (7.12% of videos). In comparison, applying the same procedure to the algorithm providing the semantic segmentation a perfect agreement was obtained for 50.87% of the videos, which increases to 83.00% when allowing a disagreement up to 1 point.

As regards the exam-based score, the level of agreement $E_{agr}(TH_i)$ as a function of TH_i is depicted in Fig. 3.4. In particular, it is observable how the boxes representing

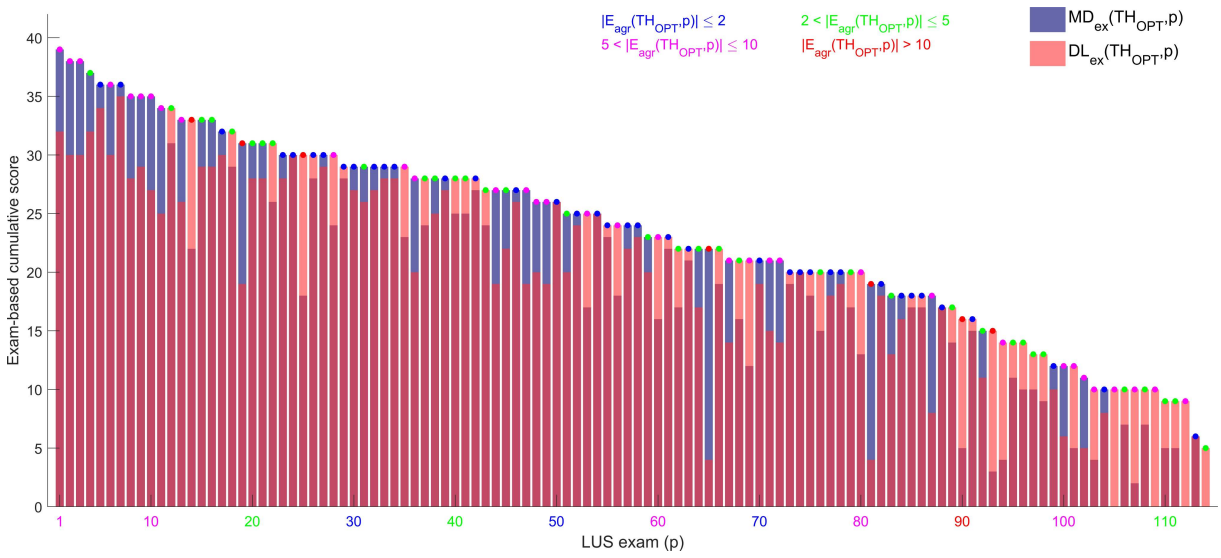

Figure 3.6: The exam-based cumulative scores for each LUS exam p are depicted (first algorithm, Section 3.2.3, applied with optimal threshold TH_{OPT}). MD exam-based scores ($MD_{ex}(p)$) and DL exam-based scores ($(DL_{ex}(TH_{OPT}, p))$) are depicted in blue and red bars, respectively. Each exam is colored in blue, green, purple, and red, depending on the disagreement interval.

$E_{agr}(TH_i)$ widen as TH_i increases when TH_i ranges from 1% to 50%. Therefore, the variability of $E_{agr}(TH_i)$ increases as TH_i increases, as well as its median values. When considering the optimal threshold $TH_{OPT} = 1\%$, the median value is close to 0 (i.e., 0.5), whereas the 25th and 75th percentiles are -3 and 4, respectively. Adopting the optimal threshold thus leads to an exam-level agreement of 33.33%, 64.03%, and 93.86% when allowing a disagreement of 2, 5, and 10 points (Fig. 3.5).

The exam-based cumulative scores for each LUS exam p [$MD_{ex}(p)$ and $DL_{ex}(TH_{OPT}, p)$] are depicted in Fig. 3.6. In the stratification between patients at high risk of clinical worsening (exam-based cumulative score > 24 [38]) and patients at low risk (exam-based cumulative score ≤ 24 [38]) the percentage of agreement between MD and DL is 85.96%.

3.3.2 Video-level score aggregating labeled and segmented frames

The video-level agreement as a function of a combination of thresholds $(TH_i, TH^{segm}_{OPT_i})$ is shown in Fig. 3.7. The overall video-level agreement $V^{all}_{agr}(TH_i, TH^{segm}_{OPT_i})$ is maximum (54.91%) when an 8% frame-level threshold and a 63% frame-level segmentation threshold ($TH_i = TH_{OPT} = 8\%$, $TH^{segm}_{OPT_i} = TH^{segm}_{OPT} = 63\%$) are applied. Also in this case, $V^{all}_{agr}(TH_i, TH^{segm}_{OPT_i})$ has a decreasing trend when evaluating values of TH_i from 1% to 100%. However, differently from the first algorithm (Section 3.2.3), the $V^{all}_{agr}(TH_i, TH^{segm}_{OPT_i})$

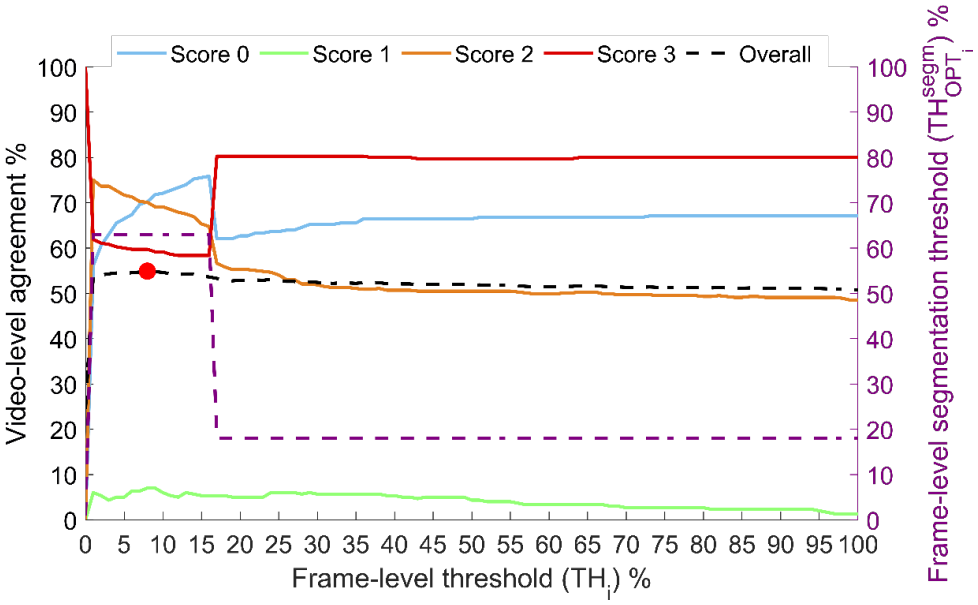

Figure 3.7: The video-level agreement as a function of threshold TH_i and TH^{segm}_i is shown (second algorithm, Section 3.2.4). The overall percentage of agreement at video-level ($V^{all}_{agr}(TH^{segm}_{OPT_i})$) is depicted with a dashed black line. The values of $TH^{segm}_{OPT_i}$ for each TH_i are depicted with a purple dashed line. The overall video-level agreement with the optimal combination of thresholds ($TH_{OPT} = 8\%$; $TH^{segm}_{OPT} = 63\%$) ($V^{all}_{agr}(TH_{OPT}, TH^{segm}_{OPT})$) is represented by a red point. The video-level agreement for scores 0, 1, 2, and 3 is indicated by light blue, light green, orange, and red lines, respectively.

values do not significantly vary (between 50.74% and 54.91%). The video-level agreement

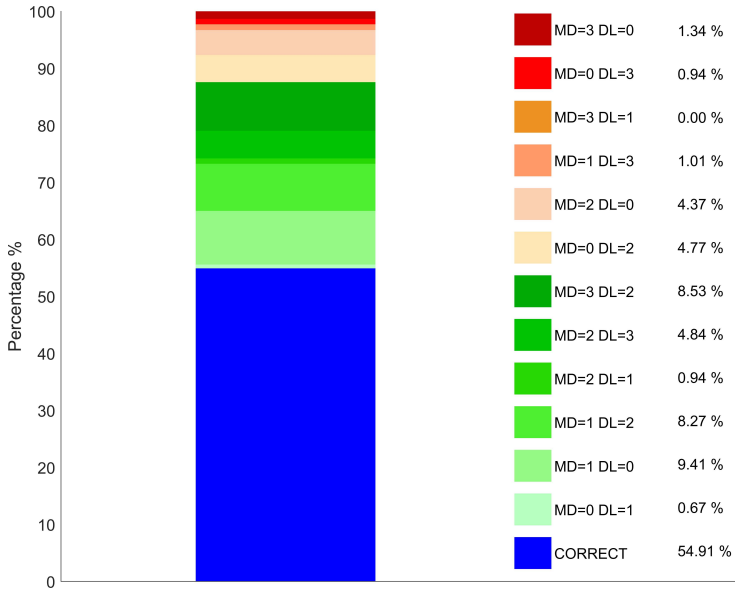

Figure 3.8: The video-level agreement and disagreement with the optimal combination of thresholds ($TH_{OPT} = 8\%$; $TH^{segm}_{OPT} = 63\%$) is shown (second algorithm, Section 3.2.4). The overall video-level agreement ($V^{all}_{agr}(TH_{OPT}, TH^{segm}_{OPT})$) is depicted in blue. The percentage of videos with disagreement between the MD scores and the DL scores is represented with different colors. A level of disagreement of 1, 2, and 3 points is represented by different tonalities of green, orange, and red, respectively.

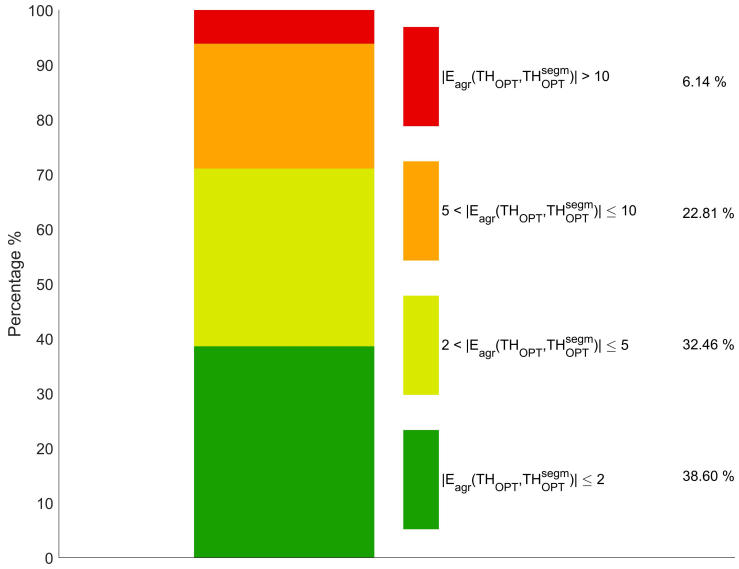

Figure 3.9: Exam-level agreement ($E_{agr}(TH_{OPT}, TH^{segm}_{OPT})$) when the optimal combination of thresholds is applied to the second algorithm (Section 3.2.4). The different colors represent the percentage of exams having a disagreement within a specific interval.

of score 3 is higher (between 58.36% and 80.00%) than the one obtained with the first algorithm (between 2.19% and 36.16%; see Fig. 3.2). In contrast, the video-level agreement of score 0 is lower (between 56.11% and 75.86%) than the one obtained with the first algorithm (between 61.44% and 99.37%; see Fig. 3.2). These different behaviors are due to the additional step applied to the videos scored as 0 or 2 by DL when using only the threshold on labeled frames (TH_i). Indeed, those videos are reprocessed by evaluating the segmented frames, which are more sensitive to score 3, thus leading to an increased number of frames labeled as 3 and a decreased number labeled as 0. The score 1 video-level agreement remains poor (between 1.34% and 7.02%), whereas, when considering TH_i ranging from 1% to 40%, the score 2 video-level agreement is lower (between 50.69% and 75.05%) than the one obtained with the first algorithm (between 51.68% and 83.56%). Figure 3.7 also shows how a significant change of the video-level agreement of scores 0 and 3 happens at $TH_i = 17\%$, i.e., when $TH^{segm}_{OPT_i}$ passes from 63% (at $TH_i = 16\%$) to 18%.

Results in Fig. 3.8 show a perfect agreement for 54.91% of the videos, and this percentage increases to 87.57% when allowing a disagreement up to 1 point. Nevertheless, for a relevant number of videos, DL tends to misclassify score 1 as score 0 (9.41% of videos), score 3 as score 2 (8.53% of videos), and score 1 as score 2 (8.27% of videos).

Figure 3.9 shows how an exam-level agreement of 38.60%, 71.06%, and 93.86% is achieved when allowing a disagreement of 2, 5, and 10 points. The exam-based cumulative scores for each LUS exam p ($MD_{ex}(p)$) and DL exam-based scores ($DL_{ex}((TH_{OPT}, TH^{segm}_{OPT}), p)$) are depicted in Fig. 3.10. It is clear from Fig. 3.10 how the second algorithm (Section 3.2.4), contrary to the first (see Fig. 3.6), tends to assign higher scores, thus over-

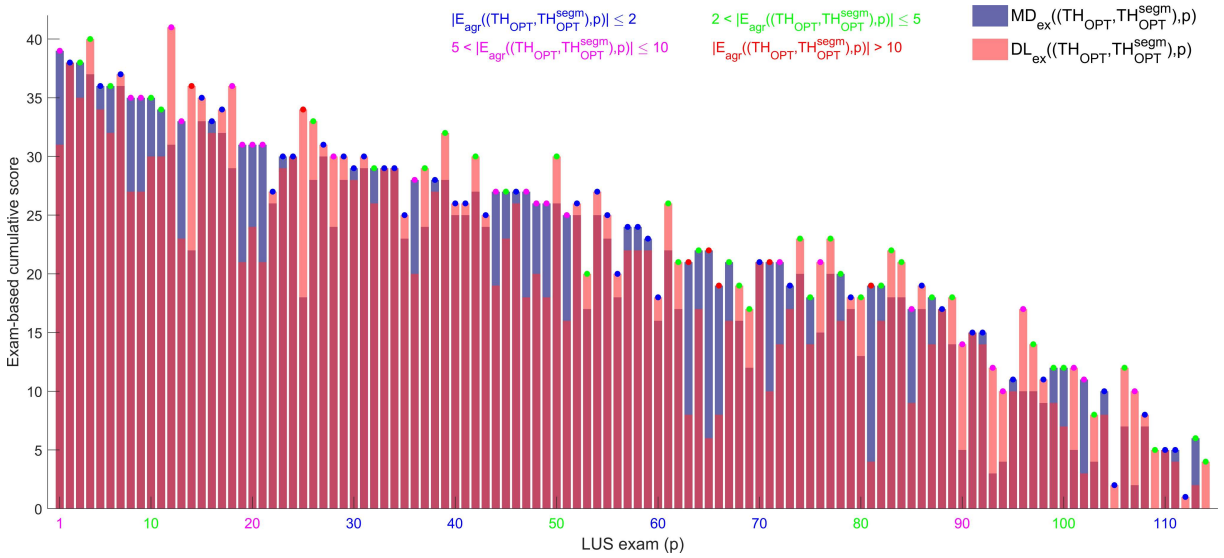

Figure 3.10: The exam-based cumulative scores for each LUS exam p are depicted (second algorithm, Section 3.2.4, applied with the optimal combination of thresholds $(TH_{OPT}, TH^{segm}_{OPT})$. MD exam-based scores ($MD_{ex}(p)$) and DL exam-based scores ($DL_{ex}((TH_{OPT}, TH^{segm}_{OPT}), p)$) are depicted in blue and red bars, respectively. Each exam is colored in blue, green, purple, and red, depending on the disagreement interval.

estimating the patient's condition. In the stratification between patients at high risk of clinical worsening and patients at low risk [38], the percentage of agreement between MD and DL is 83.33%.

As shown in Fig. 3.11, the overall video-level agreement does not significantly change when employing a different number of videos (normalized distribution of scores). In fact, even though Fig. 3.11 shows a perfect agreement for a lower percentage of videos (52.51% versus 54.91%), the percentage is very similar (87.81% versus 87.57%) when allowing a disagreement up to 1 point. Also in this case, the DL tends to misclassify, for a significant number of videos, score 1 as 0 (11.71% of videos), score 1 as 2 (10.28% of videos), and score 3 as score 2 (8.70% of videos).

The performance at video-level significantly improves when considering score 0 and score 1 as the same score (score 0/1). Specifically, as shown in Fig. 3.12, the overall video-level agreement increases to 65.59% (+10.68% compared to the performance with all four scores).

3.4 Discussion and Conclusions

Thanks to its main characteristics, i.e., safety, cost effectiveness, real-time imaging, and portability, LUS is nowadays widely adopted by clinicians to extract important informa-

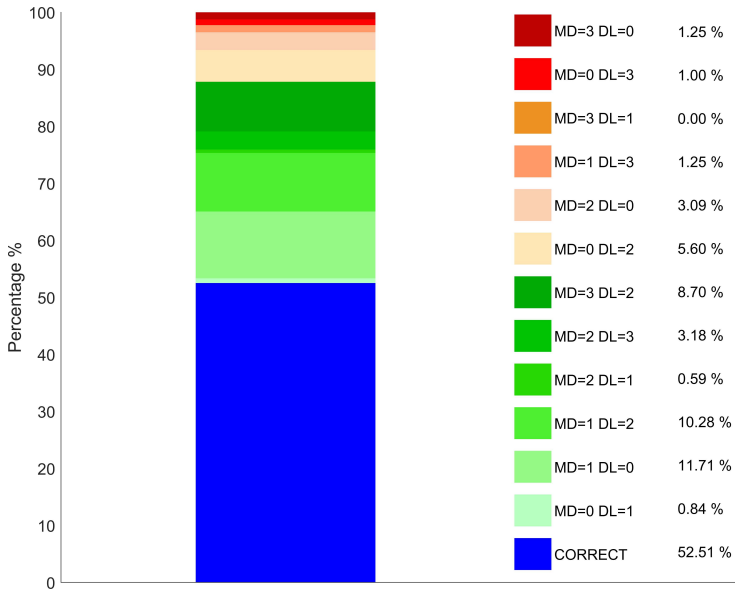

Figure 3.11: Video-level performance when the normalized distribution is analyzed (25% of videos labeled by MD for each score). The video-level agreement and disagreement with the optimal combination of thresholds ($TH_{OPT} = 8\%$; $TH^{segm}_{OPT} = 63\%$) is shown (second algorithm, Section 3.2.4). The overall video-level agreement ($V^{all}_a gr(TH_{OPT}, TH^{segm}_{OPT})$) is depicted in blue. The percentage of videos with disagreement between the MD scores and the DL scores is represented with different colors. A level of disagreement of 1, 2, and 3 points is represented by different tonalities of green, orange, and red, respectively.

tion on the state of the lung. These characteristics make LUS extremely suitable for emergency contexts [24], such as the COVID-19 pandemic [1, 32], where the monitoring and stratification of patients are fundamental.

However, to make LUS a reliable instrument for the patients' evaluation, a standardized protocol is needed. In this sense, an acquisition protocol and a scoring system, specifically designed for COVID-19 patients, have been recently proposed [1].

We acquired 1488 LUS videos from 82 patients following the above-mentioned protocol [1], and fed all the 314 879 frames to pre-trained DL algorithms [2], which provided both a frame-based classification and semantic segmentation, assigning a score from 0 to 3 [1] to each frame. Nevertheless, clinicians are interested in a video-level classification. Therefore, we evaluated possible solutions to aggregate the labeled frames to obtain a video-level score. In particular, we tested two threshold approaches (Secs. 3.2.3 and 3.2.4) and evaluate the video-level and exam-level agreement between expert clinicians (which previously scored each video) and the DL (Section 3.3).

The results show how applying a threshold (TH_i) only on labeled frames slightly outperforms an approach based only on segmented frames (perfect agreement for 51.61% and 50.87% of videos, respectively). However, when the thresholds on labeled (TH_i) and segmented (TH_i^{segm}) frames are combined, the video-level performance improves (perfect agreement for 54.91% of the videos). As regards the exam-based evaluation, the exam-level agreement is higher when applying the combined strategy (Section 3.3.2) and allowing a

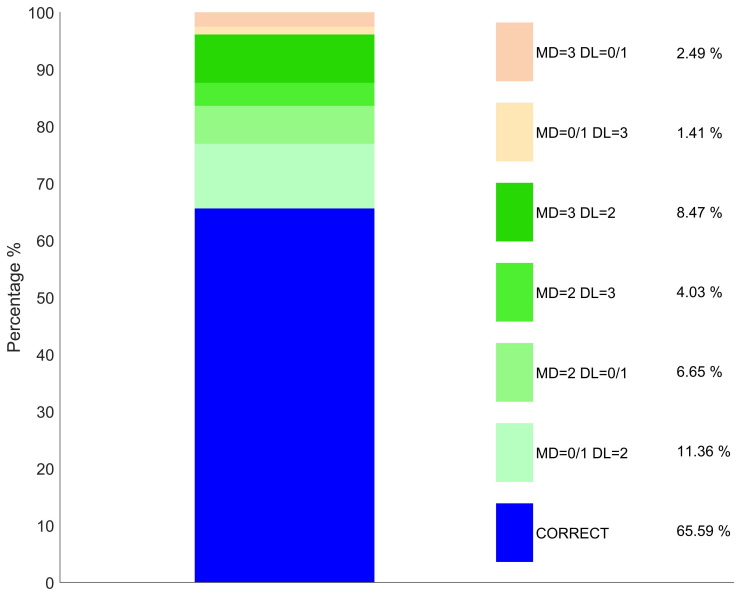

Figure 3.12: Video-level performance when scores 0 and 1 are treated as a unique score (score 0/1). The video-level agreement and disagreement with the optimal combination of thresholds ($TH_{OPT} = 8\%$; $TH_{OPT}^{segm} = 63\%$) is shown (second algorithm, Section 3.2.4). The overall video-level agreement ($(V_a^{all} gr(TH_{OPT}, TH_{OPT}^{segm}))$) is depicted in blue. The percentage of videos with disagreement between the MD scores and the DL scores is represented with different colors. A level of disagreement of 1 and 2 points is represented by different tonalities of green, and orange, respectively.

disagreement of 2 and 5 points (38.60% and 71.06% versus 33.33% and 64.03%), but is equal when allowing a disagreement of 10 points (93.86%). Moreover, when evaluating the percentage of agreement between MD and DL in the stratification between patients at high risk of clinical worsening and patients at low risk [38], the approach based only on labeled frames (Section 3.2.3) outperforms the combined approach (Section 3.2.4) of about 2.63% (85.96% versus 83.33%).

In conclusion, the combined approach generally outperforms the first approach (Section 3.2.3). However, the improvement is not significant, and the first approach seems more stable to threshold (TH_i) variations (see Fig. 3.2). Therefore, the first approach can be considered as a self-standing strategy to classify LUS videos starting from a frame-based classification. Nevertheless, the semantic segmentation remains essential for clinicians, as it provides the explainability of the decision by highlighting the specific LUS patterns.

As for future work, we aim at expanding the existing database and training the DL algorithms not on fame-based labeled data but on video-based labeled data, which is more consistent with the approach followed by clinicians when evaluating LUS videos.

Chapter 4

On the Impact of Different Lung Ultrasound Imaging Protocols in the Evaluation of Patients Affected by Coronavirus Disease 2019, How Many Acquisitions Are Needed?

This Chapter[3] *presents a preliminary study aiming at evaluating how many scanning areas should be considered during a LUS examination of COVID-19 patients. It is indeed fundamental to find the optimal trade-off between the acquisition time and accuracy. Specifically, to reduce the scanning time, a lower number of acquisition areas is preferred, but this could lead to underestimations. In this study, the results obtained with the 14-areas protocol we previously developed were compared with other protocols based on 4, 8, 10, and 12 areas. These results showed how the posterior areas are fundamental to capture the most important findings in patients with COVID-19, and how a 12-areas protocol seemed to be a better trade-off between acquisition time and accuracy.*

4.1 Introduction

The use of lung ultrasound (LUS) has widely spread during the coronavirus disease 2019 (COVID-19) pandemic [32]. Thanks to its main characteristics (real-time imaging, safety, portability, and availability), LUS can be used for stratification and monitoring of patients with COVID-19, as well as for triage of symptomatic patients. Lung ultrasound had shown its capability to detect alterations along the lung surface even before the COVID-19 pandemic. Nevertheless, being mainly based on the interpretation of imaging artifacts, e.g.,

[3]This Chapter appears in:
[J2] F. Mento, T. Perrone, V. N. Macioce, F. Tursi, D. Buonsenso, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, "On the Impact of Different Lung Ultrasound Imaging Protocols in the Evaluation of Patients Affected by Coronavirus Disease 2019: How Many Acquisitions Are Needed?," in *Journal of Ultrasound in Medicine*, vol. 40, no. 10, pp. 2235-2238, October 2021.

horizontal (A-lines) and vertical (B-lines) artifacts, LUS leads to qualitative and subjective analyses [42]. Standardization is therefore a crucial aspect, especially for defining the optimal imaging settings and the most suitable transducer. Moreover, it is fundamental to favor the reproducibility and increase the accuracy of LUS findings. In this pandemic context, a proposal for LUS standardization was recently presented, in which a 4-level score was defined, together with the acquisition time, transducers, imaging settings, and landmarks [1]. Other protocols and scanning procedures were proposed both during the COVID-19 pandemic and previously [25]. Finding the optimal trade-off between the acquisition time and accuracy represents one of the main challenges. Specifically, to reduce the scanning time, a lower number of acquisition areas is preferred, but this could lead to underestimations.

4.2 Materials and Methods

In this multicenter study, we investigated how investigating different scanning areas influences the evaluation of a patient. Specifically, we acquired LUS videos from 88 patients by following the protocol of Soldati *et al.* [1] based on 14 scanning areas per patient. A score ranging from 0 to 3 was assigned to each video, according to the 4-level scoring system [1]. Successively, each patient was classified according to the highest score (from 0 to 3) assigned to his or her 14 videos. Finally, we considered different subgroups of scanning areas to reevaluate the worst score of each patient and compared this value with the worst score obtained by the reference protocol [1]. The percentage of agreement was hence computed by summing the number of patients sharing the same worst score from the reference protocol [1] and dividing it by the total number of patients. As done by Smargiassi *et al.* [30], we first evaluated the level of agreement by separately considering only the anterior (labeled 11, 12, 13, and 14 [1]), lateral (labeled 7, 8, 9, and 10 [1]), and posterior (labeled 1, 2, 3, 4, 5, and 6 [1]) areas. Then we computed the level of agreement for the 3 protocols based on 4, 8, and 12 scanning areas [25]. We named the analyzed systems system 1 (4 areas), system 2 (8 areas), system 3 (12 areas), and system 4 (14 areas). To further investigate the specific impact of the posterior areas, we analyzed the level of agreement for 3 modified versions of system 4 (with 10 areas instead of 14) obtained by considering all of the anterior and lateral areas together with the basal posteriors (1 and 4), middle posteriors (2 and 5), or apical posteriors (3 and 6).

The patient population consisted of 88 patients with a diagnosis of COVID-19 positivity by a reverse transcription polymerase chain reaction swab test. Of the 88 patients (48 male and 40 female), 9 were examined within the Lodi General Hospital (Lodi, Italy), 29 within the Fondazione Policlinico Universitario Agostino Gemelli (Rome, Italy), and 50 within the Fondazione Policlinico San Matteo (Pavia, Italy). This study was part of a protocol that has been registered (NCT04322487) and received approval from the Ethical Committee of the Fondazione Policlinico Universitario Agostino Gemelli, Istituto di Ricovero e Cura a Carattere Scientifico (protocol 0015884/20 ID 3117), of Milano area 1, the Azienda Socio-Sanitaria Territoriale Fatebenefratelli-Sacco (protocol N0031981), and the Fondazione Policlinico Universitario San Matteo (protocol 20200063198). All patients gave informed consent. The patients' ages ranged from 26 to 95 years (average, 59.9

years).

4.3 Results and Conclusions

Figure 4.1 shows the distributions of scores for anterior, lateral, and posterior areas. The highest percentage of score 0 (35.23%) was observed in the anterior areas. In contrast, posterior areas showed the highest percentage of score 3 (30.87%) compared to anterior (10.80%) and lateral (16.76%) areas. Jointly considering score 2 and score 3, the overall percentages for the anterior, lateral, and posterior areas were 38.35%, 54.26%, and 58.52%, respectively. Therefore, the highest scores were focused on the lateral and posterior areas, whereas the lowest scores were focused on the anterior area. As a consequence, the levels of agreement with system 4 for just the anterior, lateral, and posterior areas were 40%, 58%, and 92% (Figure 4.1). These results were consistent with the results achieved by Smargiassi *et al.* [30] and highlight the importance of the posterior areas.

Figure 4.2, top, shows how the distributions of scores vary with different systems: i.e., system 1 (scanned areas, 7, 9, 12, and 14 [1]), system 2 (scanned areas, 7–14 [1]), system 3 (scanned areas, 1, 3, 4, 6, and 7–14 [1]), and system 4 (scanned areas, 1–14 [1]). Of particular interest is the distribution of score 3, whose percentage values were 14.78%, 14.20%, 19.90%, and 21.72% for systems 1, 2, 3, and 4, respectively. Consequently, the levels of agreement with system 4 for systems 1, 2, and 3 were 57%, 70%, 99% (Figure 4.2, top). This result suggests that the 12-areas scanning protocol [25] leads to performance comparable with that of the reference protocol (14 areas) [1]. Figure 4.2, bottom, shows

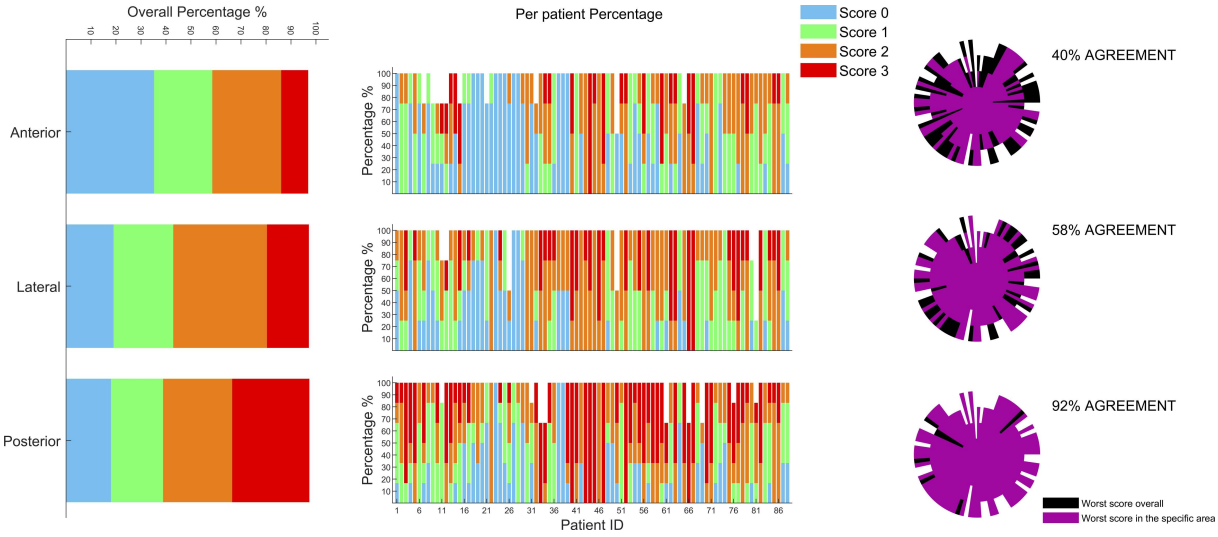

Figure 4.1: The overall distribution of assigned scores divided per specific area (anterior, lateral, and posterior) is depicted on the left side. The percentage of scores assigned for each area and for each patient is depicted in the center. The level of agreement for all of the patients' scanning in the different areas is shown on the right side (For further details about the structure of agreement graphs, see Smargiassi *et al* [30].) Each patient was classified according to the worst score. The reference system is system 4.

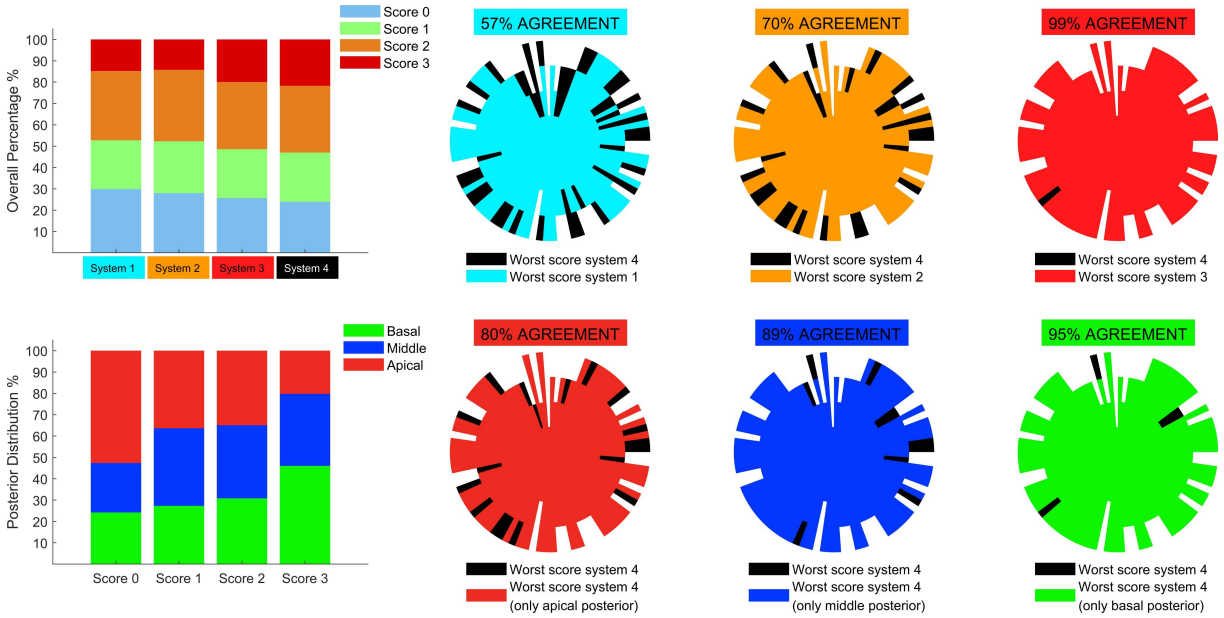

Figure 4.2: The overall distribution of score considering the 4 systems is depicted on the top left. The level of agreement between systems 1, 2, and 3 with respect to system 4 is also shown on the top. (For further details about the structure of agreement graphs, see Smargiassi *et al* [30].) The distribution of each score in the posterior areas (basal, middle, and apical) is also depicted on the bottom left. The level of agreement between the 3 modified versions of system 4 (10 zones instead of 14: i.e., all of the anterior and lateral areas together with apical posteriors, middle posteriors, or basal posteriors) with respect to system 4 is also shown on the top.

the score distributions for the posterior areas. The highest percentage of score 0 (52.63%) was observed when scanning the apical posterior areas, whereas the highest percentage of score 3 (46.01%) was observed when scanning the basal posterior areas. Consequently, the levels of agreement between system 4 and the modified systems [25] were 80%, 89%, and 95% when the scanned posterior areas were the apical, middle, and basal, respectively. This highlights the importance of scanning basal posterior areas.

To conclude, the results show that the posterior areas are fundamental to capture the most important findings in patients with COVID-19. Moreover, the 12-areas system seems to be a better trade-off between the acquisition time and accuracy. However, the 10-areas system also seems to be sufficiently accurate, as long as the basal posteriors are included.

Chapter 5

Lung Ultrasound in COVID-19 and Post-COVID-19 Patients, an Evidence-Based Approach

This Chapter[4] *presents an extensive study analyzing the impact of different LUS imaging protocols on the evaluation of COVID-19 and post-COVID-19 LUS data. Specifically, LUS data from 220 patients were collected, 100 COVID-19 positive and 120 post-COVID-19. A validated and standardized imaging protocol based on 14 scanning areas and a 4-level scoring system was used. This dataset was exploited to compare the capability of 5 imaging protocols, respectively based on 4, 8, 10, 12, and 14 scanning areas, to intercept the most important LUS findings. This to evaluate the optimal trade-off between a time-efficient imaging protocol and an accurate LUS examination. We also performed a longitudinal study, aimed at investigating how to eventually simplify the protocol during follow-up. Additionally, we present results on the agreement between AI models and LUS experts with respect to LUS data evaluation. A 12-areas protocol emerges as the optimal trade-off, for both COVID-19 and post-COVID-19 patients. For what concerns follow-up studies, it appears not to be possible to reduce the number of scanning areas. Finally, COVID-19 and post-COVID-19 LUS data seem to show differences capable to confuse AI models that were not trained on post-COVID-19 data, supporting the hypothesis of the existence of LUS patterns specific to post-COVID-19 patients.*

5.1 Introduction

During the recent coronavirus disease 2019 (COVID-19) pandemic, lung ultrasound (LUS) has emerged as a powerful ally for clinicians. In fact, thanks to ultrasound technologies portability, cost-effectiveness, and safety, LUS has been utilized extensively around the

[4]This Chapter appears in:
[J3] L. Demi, F. Mento, A. Di Sabatino, A. Fiengo, U. Sabatini, V. N. Macioce, M. Robol, F. Tursi, C. Sofia, C. Di Cienzo, A. Smargiassi, R. Inchingolo, T. Perrone, "Lung Ultrasound in COVID-19 and Post-COVID-19 Patients, an Evidence-Based Approach," in *Journal of Ultrasound in Medicine*, Online ahead of print, December 2021.

world to assess the condition of the lung in patients suspected or affected by COVID-19 [1, 32, 110–116]. Specifically, LUS has been utilized to intercept the presence of COVID-19-associated interstitial pneumonia, and monitor its evolution. To this end, a variety of imaging protocols and scoring systems have been proposed in the literature [25]. However, fundamental aspects such as the amount and spatial distribution of areas of the chest to be scanned are often defined arbitrarily and not following an evidence-based approach. Defining the right amount and distribution of scanning areas is of significant importance for LUS, given that ultrasound imaging can only provide local information on the status of the lung surface. Consequently, reducing the scanning areas in order to simplify the examination does impact directly on the extent of the inspected lung surface. This is particularly relevant for COVID-19, given the patchy distribution of the relevant findings [30, 117–121]. Moreover, extended studies on LUS findings on post-COVID-19 patients are currently lacking. Therefore, in this multicenter study we investigate the impact of the amount and distribution of scanning areas on the accuracy of the LUS examination. To this end, we analyzed LUS data acquired on a population of 220 patients. Specifically, a 14-areas acquisition protocol and a 4-level scoring system were utilized. The prognostic value of this approach has been investigated through a study conducted at the Fondazione Policlinico San Matteo (Pavia, Italy), and involving 52 patients [38]. Results showed how patients showing a cumulative LUS score (the sum of the scores over the 14 areas scanned) higher than 24 had an almost 6-fold increase in the odds of worsening. Moreover, we investigated LUS findings variability with respect to the implemented imaging protocol. Five imaging protocols were considered, respectively based on 4, 8, 10, 12, and 14 scanning areas. This approach allows to define the optimal trade-off between a simple and time-efficient LUS evaluation (which requires minimizing the number of areas to be scanned) and an accurate LUS examination (which requires maximizing the areas to be scanned).

Of the 220 patients, 100 were COVID-19 positive at the time they were scanned, while 120 patients were post-COVID-19, that is, they were negative to reverse transcription polymerase chain reaction (RT-PCR) test after being originally diagnosed with COVID-19 by means of the same test.

To our knowledge an extensive study on LUS patterns on post-COVID-19 patients represents by itself a significant novelty to the existing scientific literature. Additionally, we report on results from a longitudinal study on a subgroup of 29 patients. These results are important to investigate the evolution of the lung condition over time, and to verify whether a simplified scanning procedure could be adopted during patients' follow-up. In conclusion, we also report on the level of prognostic agreement between LUS experts and recently developed AI algorithms [2], which were trained at implementing the previously introduced scoring system [1]. Specifically, we investigate the performance of the AI, differentiating between COVID-19 and post-COVID-19 data. The AI algorithm discussed in this manuscript was the first DL algorithm to be developed worldwide for the analysis of LUS data from COVID-19 patients. A detailed technical description of the algorithm can be found in a recently published article [2]. This algorithm was then validated in a multicenter study involving 314,879 images from 82 patients. In that study, the DL performance at scoring LUS videos was compared with that of clinical experts [45]. To our knowledge, this is the only DL algorithm that has had a similar validation

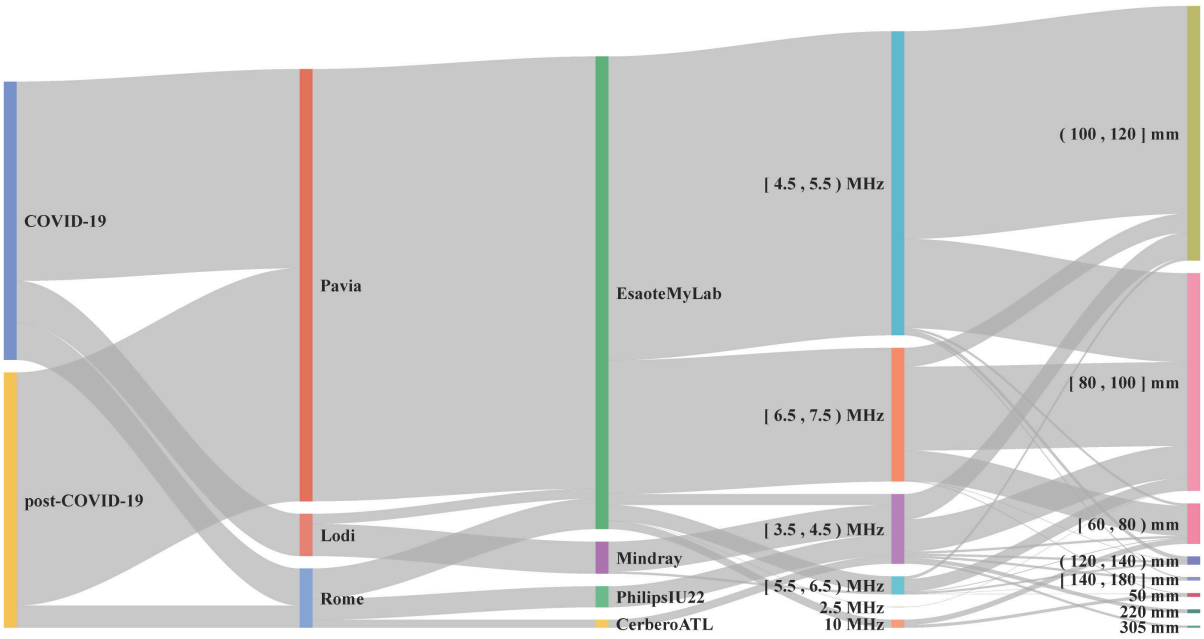

Figure 5.1: Sankey diagram illustrating the distribution of the dataset characteristics. Square and round brackets are respectively utilized to indicate whether the interval includes or not the endpoints. Data are grouped (from left to right) based on they being from COVID-19 or post-COVID-19 patients, based on the hospital where the data have been collected, on the utilized ultrasound scanner, on the imaging frequency and imaging depth. Frequencies are expressed in Hertz (MHz = 10^6 Hz) and depths in meters (mm = 10^{-3} m).

(distinguishing frame-level, video-level, and exam-level performance) for the analysis of LUS data from COVID-19 patients. Results from the multicenter study showed a level of agreement between DL and clinical experts of 85.96% in the stratification between patients at high risk of clinical worsening and patients at low risk. In this new work, we have further extended this validation to data acquired from 220 patients, and distinguished between the performance obtained for COVID-19 and post-COVID-19 patients.

The paper is organized as follows. Firstly, we present the study design and population, then we describe the utilized LUS protocol. Successively, we present the methods used to assess the impact of different scanning areas on the LUS exam's evaluation, and describe the design of the longitudinal study and the methods implemented for the analysis of the prognostic agreement between LUS experts and AI. Next, the results are introduced. Finally, we present the discussion and the conclusion.

5.2 Materials and Methods

5.2.1 Study Design and Population

The studied population consists of 220 patients (138 male, 82 female, with ages ranging from 23 to 95 years, and average age equal to 63.0 years). Inclusion criteria were age

18 years or older, confirmed COVID-19 infection based on the detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on a reverse transcriptase polymerase chain reaction from a nasopharyngeal swab or bronchoalveolar wash, and a collaborative status allowing them to express informed consent. Patients were excluded if they were not able to express their consent, if they were severely obese (body mass index [BMI] $>$ 35 kg/m^2), or if they were affected by heart failure or interstitial lung disease, such as usual interstitial pneumonia or lung fibrosis secondary to rheumatologic disease. Patients' enrolment was performed, for the acute COVID-19 patients, at the internal medicine ward (converted in a sub-intensive COVID ward) for San Matteo and Lodi General Hospital, while post-COVID-19 patients from San Matteo Hospital were outpatients. From Gemelli Hospital, acute COVID-19 patients were enrolled when hospitalized in dedicated wards converted in a sub-intensive COVID ward, while post-COVID-19 inpatients were enrolled in the pulmonology ward.

Of the 220 patients, 100 were diagnosed as COVID-19 positive by a RT-PCR swab test, and 120 are post-COVID-19 patients (mean days between last positive RT-PCR swab test and LUS examination equal to 47.85 $\pm$ 12.82). Of the 100 COVID-19 patients, 63 (35 male, 28 female, with ages ranging from 26 to 92 years, and average age equal to 63.72 years) were examined within the Fondazione Policlinico San Matteo (Pavia, Italy), 19 (16

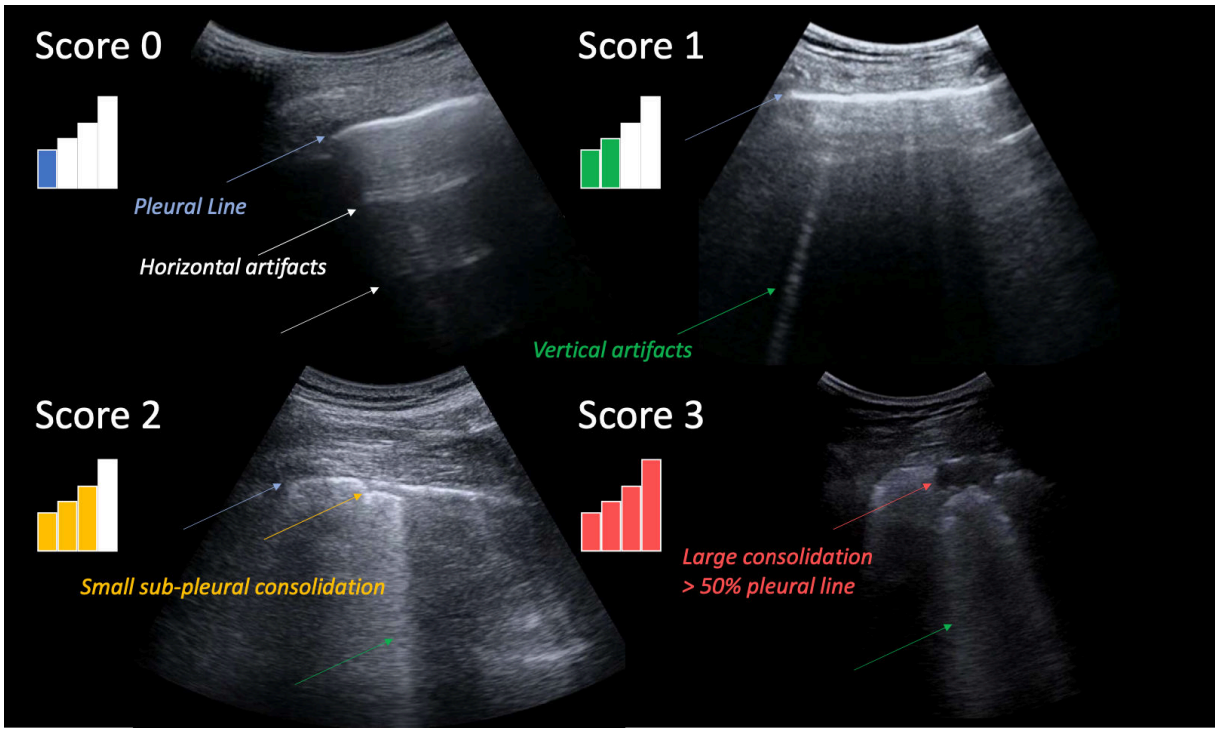

Figure 5.2: Typical LUS image associated with each level of the scoring system. A higher score is associated with a higher level of deaeration of the lung surface explored by ultrasound. A higher score is thus intended to signal a worsening of the status of the lung surface [1]. Relevant patterns are indicated by color-coded arrows. The displayed images were acquired with a convex probe.

male, 3 female, with ages ranging from 34 to 84 years, and average age equal to 63.95 years) within the Lodi General Hospital (Lodi, Italy), and 18 (8 male, 10 female, with ages ranging from 23 to 95 years, and average age equal to 52.11 years) within the Fondazione Policlinico Universitario Agostino Gemelli (Rome, Italy). Of the 120 post-COVID-19 patients, 109 (71 males, 38 females, with ages ranging from 36 to 87 years, and average age equal to 63.20 years) were examined within the Fondazione Policlinico San Matteo, and 11 (8 males, 3 females, with ages ranging from 52 to 89 years, and average age equal to 73.64 years) within the Fondazione Policlinico Universitario Agostino Gemelli. It is important to highlight how the post-COVID-19 patients examined at Pavia were scanned during follow-up and were not hospitalized at the date of LUS exam, whereas the post-COVID-19 patients examined at Rome were still hospitalized at the date of LUS exam. As a subgroup of COVID-19 patients was examined multiple times (on different dates), a total of 253 LUS exams were performed (COVID-19 positive: 94 at Pavia, 20 at Lodi, 19 at Rome; post-COVID-19: 109 at Pavia, 11 at Rome), and 3,481 LUS videos acquired (COVID-19 positive: 1,291 from Pavia, 276 from Lodi, 242 from Rome; post-COVID-19: 1,526 from Pavia, 146 from Rome), consisting of 772,780 frames (COVID-19 positive: 293,194 from Pavia, 44,288 from Lodi, 29,070 from Rome; post-COVID-19: 371,168 from Pavia, 35,060 from Rome). LUS data were acquired by LUS experts with more than 10 years of experience. Andrea Smargiassi, Riccardo Inchingolo, Tiziano Perrone and Francesco Tursi respectively acquired the data collected at the Gemelli, San Matteo and Lodi Hospital.

The data from Pavia have been acquired using a convex probe with an Esaote MyLab Twice scanner, and an Esaote MyLab 50, setting an imaging depth from 5 to 13 cm (depending on the patient) and an imaging frequency from 2.5 to 6.6 MHz (depending on the scanner). The data from Lodi have been acquired using a convex probe with an Esaote Mylab Sigma scanner, and a MindRay TE7, setting an imaging depth from 8 to 12 cm (depending on the patient) and an imaging frequency from 3.5 to 5.5 MHz. The data from Rome have been acquired using both convex and linear probes with an Esaote MyLab 50, an Esaote MyLab Alpha, a Philips IU22, and an ATL Cerbero, setting an imaging depth from 5 to 30 cm (depending on the patient) and an imaging frequency from 3.5 to 10 MHz (depending on the scanner).

Figure 5.1 shows a Sankey diagram where the distribution of the dataset characteristics is illustrated in detail. As visible, the majority of the data have been acquired with an imaging frequency ranging from 2.5 to 7.5 MHz and an imaging depth from 8 to 12 cm.

This study was part of a protocol that has been registered (NCT04322487) and received approval from the Ethical Committee of the Fondazione Policlinico Universitario San Matteo (protocol 20200063198), of the Fondazione Policlinico Universitario Agostino Gemelli, Istituto di Ricovero e Cura a Carattere Scientifico (protocol 0015884/20 ID 3117), of Milano area 1, the Azienda Socio-Sanitaria Territoriale Fatebenefratelli-Sacco (protocol N0031981). All patients gave informed consent.

5.2.2 LUS Acquisition Protocol

All patients were examined following the standardized acquisition protocol presented by Soldati *et al.* and based on 14 scanning areas [1]. According to this protocol, a score

ranging from 0 to 3 was assigned to each video by LUS experts (TP, AS, and FT) [1]. Figure 5.2 shows an example of LUS image for each level of the scoring system. The 4 levels are defined based on the current understanding of ultrasound interaction with lung tissue. Score 0 corresponds to a continuous pleural line with associated horizontal artifacts. These artifacts are generally referred to as A-lines, and are due to the high reflectivity of the non-pathological lung surface preventing ultrasound waves to propagate beyond the pleural line. Ultrasound waves are thus scattered multiple times between the lung surface and the probe, giving rise to this particular horizontal pattern. Score 1 signals instead the appearance of the first abnormalities. The pleural line is not continuous anymore and vertical artifacts are visible. We prefer to adopt the general term vertical artifact over a different term generally found in LUS literature (i.e., B-lines). This choice is motivated as to avoid the ambiguity related to the definition of B-lines. Moreover, the presence, and not the number, of vertical artifacts is considered. In fact, recent clinical studies [58, 59] showed how the visualization of vertical artifacts is strongly influenced by key imaging settings such as the imaging frequency and bandwidth. Moreover, it is also very important to stress how vertical artifacts are not specific to COVID-19, and simply signal the presence of local alterations along the lung surface. Their appearance during an ultrasound exam is considered to be due to the formation, along the lung surface, of channels accessible to ultrasound, which can indeed be generated in many pathological states of the lung once volumes originally filled by air are occupied by media that are acoustically much more similar to the intercostal tissue (e.g., water, blood, and tissue) [58, 59]. Score 2 is associated with the appearance of small-to-large consolidated areas. Differently that with horizontal and vertical artifacts, consolidations are not artifacts, but anatomical findings that appear as hypoechoic areas (darker areas) along the lung surface. The loss of echogenicity of the consolidated areas is a reflection of the loss of aeration and signal the transition of these areas toward acoustic properties similar to soft tissue. They thus signal deaeration. Below the consolidations, vertical artifacts are generally found. The latter are most likely associated with the presence of areas not yet fully deaerated. Ultimately, score 3 is associated with the presence of large, extended (>50% of the pleural line) vertical artifacts (sometimes referred to White Lung in the literature), with or without large consolidations.

5.2.3 Impact of Different Scanning Areas on Exam's Evaluation

Consistently with a previous study [28], we classified each exam according to the highest score (from 0 to 3) assigned to the corresponding 14 LUS videos. Then, different subgroups of scanning areas were considered to reevaluate the worst score of each exam and compare the obtained value with the worst score obtained by the reference protocol (14 scanning areas) [1, 28]. Finally, we computed the percentage of agreement by summing the number of exams sharing the same worst score from the reference protocol (named system 4) [1] and dividing it by the total number of exams [28]. Given the presence of two significantly different groups (i.e., COVID-19 patients and post-COVID-19 patients), this evaluation was separately performed for each group.

Firstly, we analyzed the level of agreement by separately considering only the anterior (named 11, 12, 13, 14 [1]), lateral (named 7, 8, 9, 10 [1]), and posterior (named 1, 2, 3,

4, 5, and 6 [1]) areas [28, 30]. Secondly, we evaluated the level of agreement for three different protocols based on 4 (named system 1), 8 (named system 2), and 12 (named system 3) scanning areas [25, 28]. Specifically, system 1 is based on scanning areas 7, 9, 12, and 14, system 2 on scanning areas from 7 to 14, system 3 on scanning areas 1, 3, 4, 6 and from 7 to 14, whereas the reference system (system 4) is based on all the scanning areas (from 1 to 14) [1, 28]. Moreover, given the impact of posterior areas in the exam's

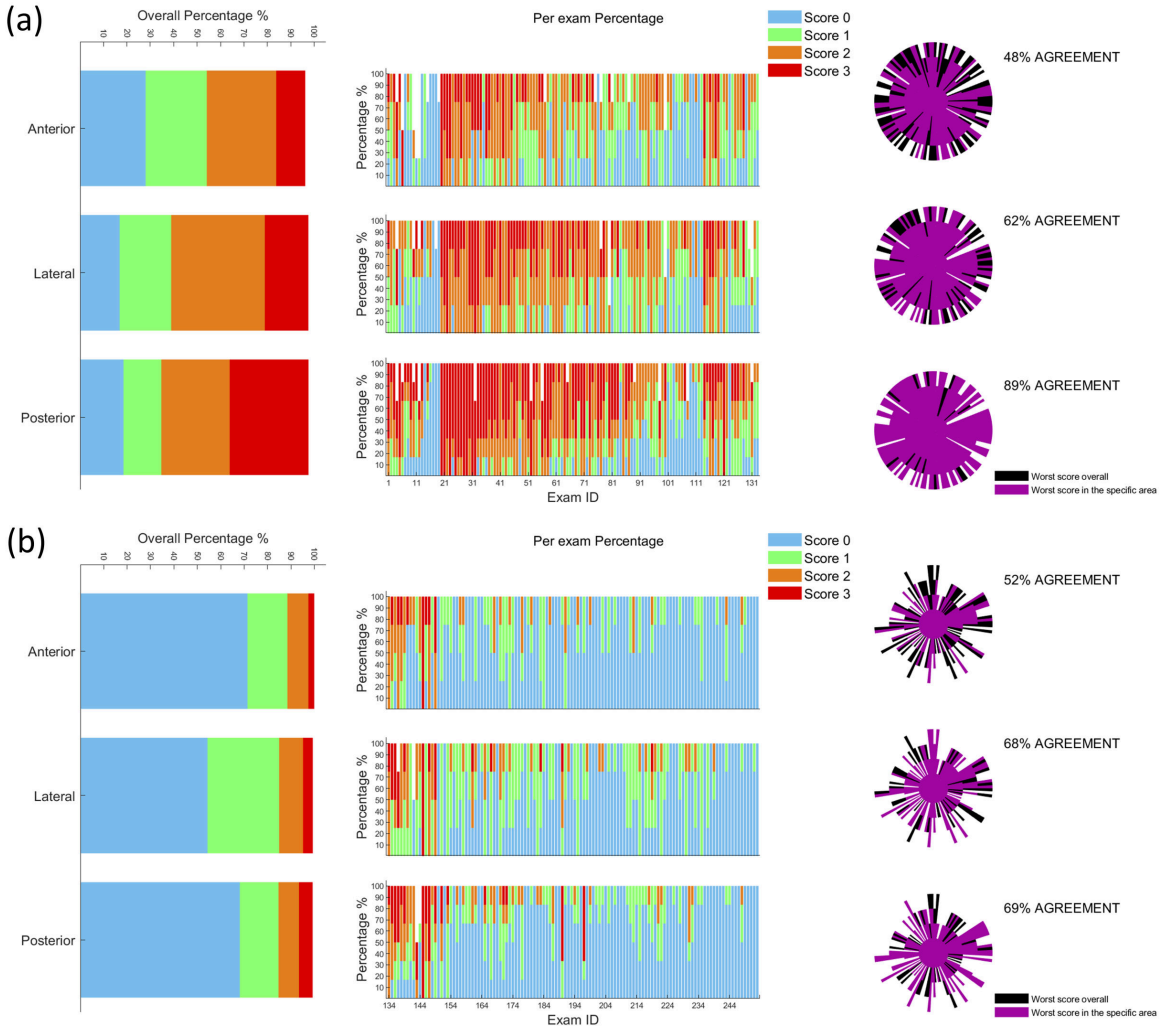

Figure 5.3: A, Graphs referring to LUS exams performed on COVID-19 patients; B, graphs referring to LUS exams performed on post-COVID-19 patients. The overall distributions of scores, divided per specific area (anterior, lateral, and posterior), are depicted on the left. The percentage of scores assigned for each area and for each exam is depicted in the center. The level of agreement is shown on the right. Each exam is represented by a beam of the polar plot. The score is indicated by the length of a beam. The longer the beam, the higher the score. For further details about the structure of agreement graphs see Smargiassi *et al* [30]. Each exam was classified according to the worst score. The reference system is system 4 (14 scanning areas).

evaluation [28, 30]., three modified versions of system 4 (based on 10 areas instead of 14) were evaluated [28]. In particular, these three modified versions were obtained by considering all the anterior and lateral scanning areas (from 7 to 14) together with the basal posteriors (1 and 4), middle posteriors (2 and 5), or apical posteriors (3 and 6) [28].

5.2.4 Longitudinal Study

We performed a longitudinal study with a subgroup of 29 COVID-19-positive patients (15 males, 14 females, with ages ranging from 39 to 92 years, and average age equal to 67.55 years) that underwent LUS exams twice (in different dates; days between the first and second LUS exams equal to 6.93 ± 5.44), to evaluate how the score assigned to each area of system 4 changes between the two exams. Specifically, we computed, for each patient, the difference between the score assigned to each scanning area at the first LUS exam and at the second LUS exam, which will be referred to as Δ. Therefore, the values of Δ range from -3 to 3, where a negative value represents a worsening of the patient in the considered scanning area, whereas a positive value represents an improvement of the patient in that scanning area. Then, we computed, for each patient, the mean value of Δ by averaging the Δ values obtained for each scanning area, and the minimum and maximum values of Δ (similarly obtained). Hence, we obtained an error bar for each patient, where its length is given by the distance between the minimum Δ value and the maximum Δ value. A long error bar with a mean value in the middle would represent a heterogeneous change of Δ with respect to the different scanning areas, thus highlighting the necessity to scan all the 14 areas every time a new LUS exam is required. In contrast, a short error bar would represent a homogeneous change of Δ with respect to the different scanning areas, thus suggesting the possibility to scan only a subgroup of areas and implicitly predict the scores of the other areas.

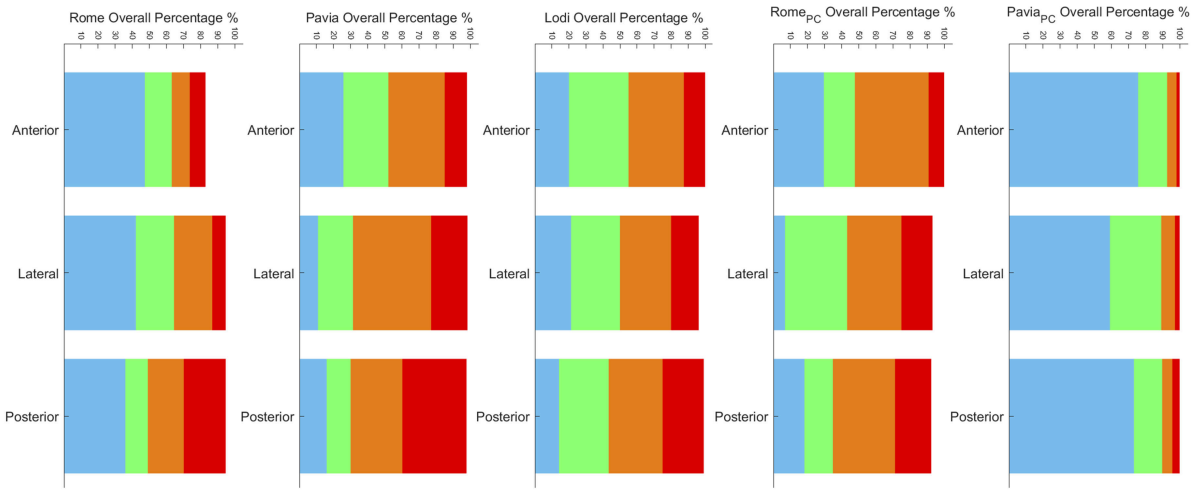

Figure 5.4: The overall distributions of scores, divided per specific area (anterior, lateral, and posterior) and per each subgroup (from left to the right: COVID-19 patients of Rome, Pavia, and Lodi, and post-COVID-19 patients of Rome and Pavia).

5.2.5 Prognostic Evaluation

The prognostic value of the reference acquisition protocol and scoring system [1] has been recently evaluated based on the cumulative score, that is, the sum of the scores on each of the 14 scanning areas [38]. As the score for each LUS video ranges from 0 to 3, the cumulative score ranges from 0 to 42. Specifically, when the exam-based cumulative score (also called sum of scores) is greater than 24, the patient is considered at high risk of clinical worsening, whereas, when the exam-based cumulative score is less than or equal to 24, the patient is considered at low risk. This threshold follows from the results of a study conducted at the Fondazione Policlinico San Matteo (Pavia, Italy), and involving 52 patients [38]. This strategy can therefore help the stratification between patients at high risk of clinical worsening and patients at low risk [38].

In this part of the study, we assessed the capability of recently developed artificial intelligence (AI) algorithms [2] of automatically stratifying patients at high risk of clinical worsening from patients at low risk. We hence compared, for each LUS exam, the cumulative score values obtained from the analysis of the data performed by LUS experts, with that provided by the AI [45]. Specifically, we considered clinicians and AI in prognostic agreement when both the cumulative scores are greater than 24 (high risk of worsening) or less than or equal to 24 (low risk of worsening). To perform this comparison, we needed to classify each video with a single score. However, the AI provided frame-level labeling. Hence, we used an aggregation technique consisting of assigning to each video the highest score assigned at least at a given percentage of frames (threshold) composing the video [45]. In this work we applied the optimal threshold (1%) [45]. This threshold was derived from the analysis of a dataset obtained within a multicenter study, and involving 314,879 images from 82 patients. The technical details of the implementation are described in a previous publication [45].

5.3 Results

5.3.1 Impact of Different Scanning Areas on Exam's Evaluation

Figure 5.3 shows the score distributions for anterior, lateral, and posterior areas, for LUS exams performed on COVID-19 patients (Figure 5.3a) and post-COVID-19 patients (Figure 5.3b). Considering COVID-19 patients (Figure 5.3a), the highest percentage of score 0 (28.01%) was observed in the anterior areas, whereas posterior areas show the highest percentage of score 3 (33.71%). Jointly considering score 2 and 3, the percentages for the anterior, lateral, and posterior areas are 42.11, 58.65, and 62.91%, respectively. This result highlights how the highest scores are focused on the lateral and posterior areas. Consequently, the levels of agreement with system 4 for just the anterior, lateral, and posterior areas are 48, 62, and 89%, respectively (Figure 5.3a). All these results on COVID-19 patients are consistent with the results achieved by Mento *et al* [28]. On the other hand, Figure 5.3b shows how the distributions of scores are different in post-COVID-19 patients. Specifically, score 0 is significantly more present (71.46, 54.37, and 68.19% for anterior, lateral, and posterior areas), whereas score 3 is the less frequent (2.50, 4.17, and 5.83% for anterior, lateral, and posterior areas). Given the almost homogeneous

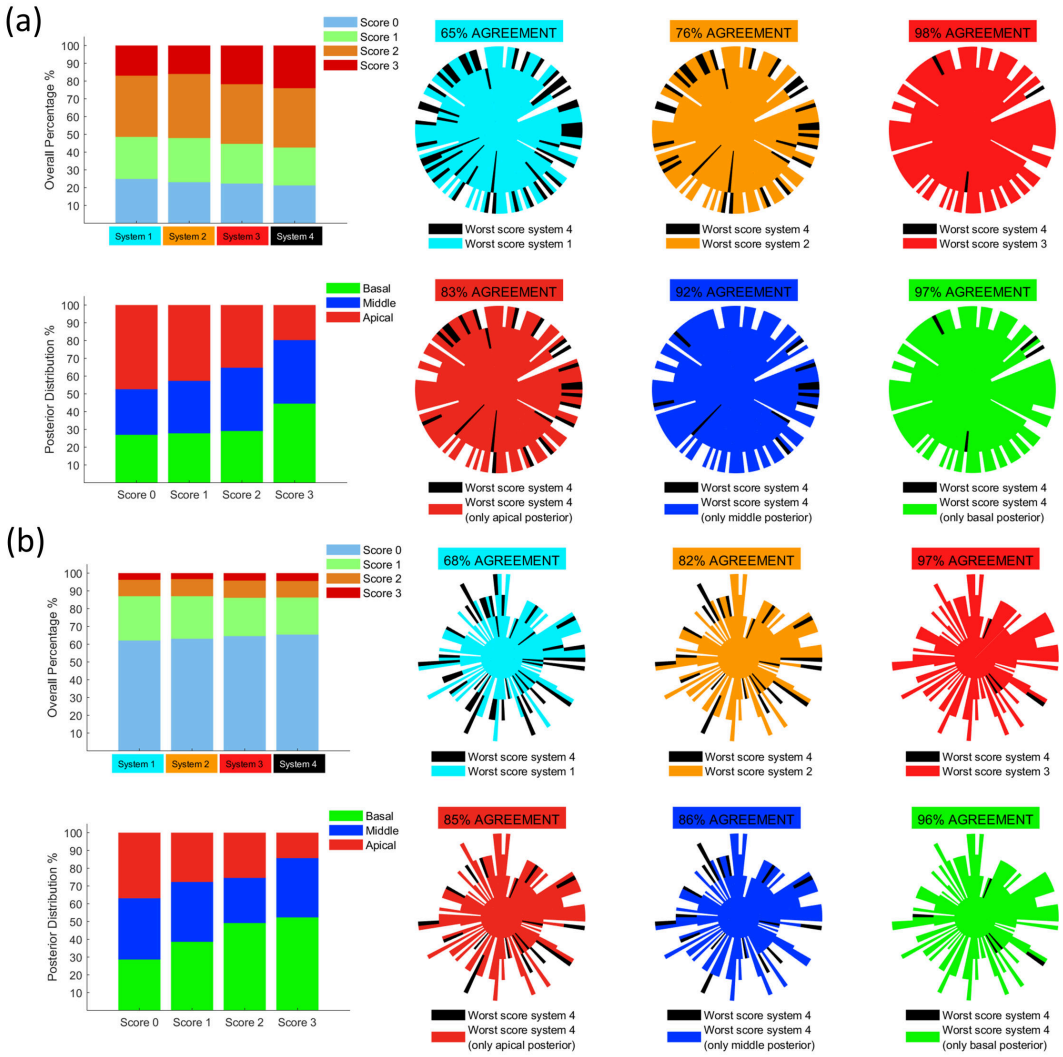

Figure 5.5: A, Graphs referring to LUS exams performed on COVID-19 patients; B, graphs referring to LUS exams performed on post-COVID-19 patients. On the top left of (A) and (B) the overall distributions of scores considering the four systems are shown, and, on the top right of (A) and (B), the level of agreement between systems 1, 2, and 3 with respect to system 4 is depicted. Each exam is represented by a beam of the polar plot. The score is indicated by the length of a beam. The longer the beam, the higher the score. For further details about the structure of agreement graphs see Smargiassi *et al* [30]. On the bottom left of (A) and (B) the distributions of each score in the posterior areas (basal, middle, and apical) are shown, and, on the bottom right, the level of agreement between the 3 modified versions of system 4 (10 zones instead of 14: i.e., all of the anterior and lateral areas together with apical posteriors, middle posteriors, or basal posteriors) with respect to system 4 is shown.

distribution of the worst scores (i.e., scores 2 and 3), for post-COVID-19 patients, the levels of agreement with system 4 for just the anterior, lateral, and posterior areas are 52, 68, and 69%, respectively (Figure 5.3b).

Figure 5.4 shows the overall distributions of scores divided per specific area and per each subgroup (i.e., acquisition center and kind of patients). It is interesting how the distributions of score 0 and score 3 in COVID-19 patients are consistent among the three acquisition centers. Moreover, it is clear how in post-COVID-19 patients the percentage of worst scores (i.e., scores 2 and 3) is significantly higher when comparing patients that were still hospitalized at the date of LUS exam (Rome) with patients that were not (Pavia) (57.57% vs 10.24% in the posterior areas).

Figure 5.5 shows how the distributions of scores vary with different systems, for both COVID-19 (Figure 5.5-a) and post-COVID-19 (Figure 5.5b) patients. As introduced in section 5.2, five systems have been investigated, that is, system 1 (scanned areas, 7, 9, 12, and 14), system 2 (scanned areas, 7–14), system 3 (scanned areas, 1, 3, 4, 6, and 7–14), system 4 (scanned areas, 1–14), and a system based on 10 scanning areas (scanned areas 7–14 plus 2 posterior areas). Figure 5.5 shows how the trend of score distributions and percentage of agreement is similar when evaluating COVID-19 patients (Figure 5.5a) and post-COVID-19 patients (Figure 5.5b). In fact, even though the score distributions of these two groups are significantly different, the levels of agreement with system 4 for systems 1, 2, and 3 are 65, 76, and 98% for COVID-19 patients (Figure 5.5a, top right),

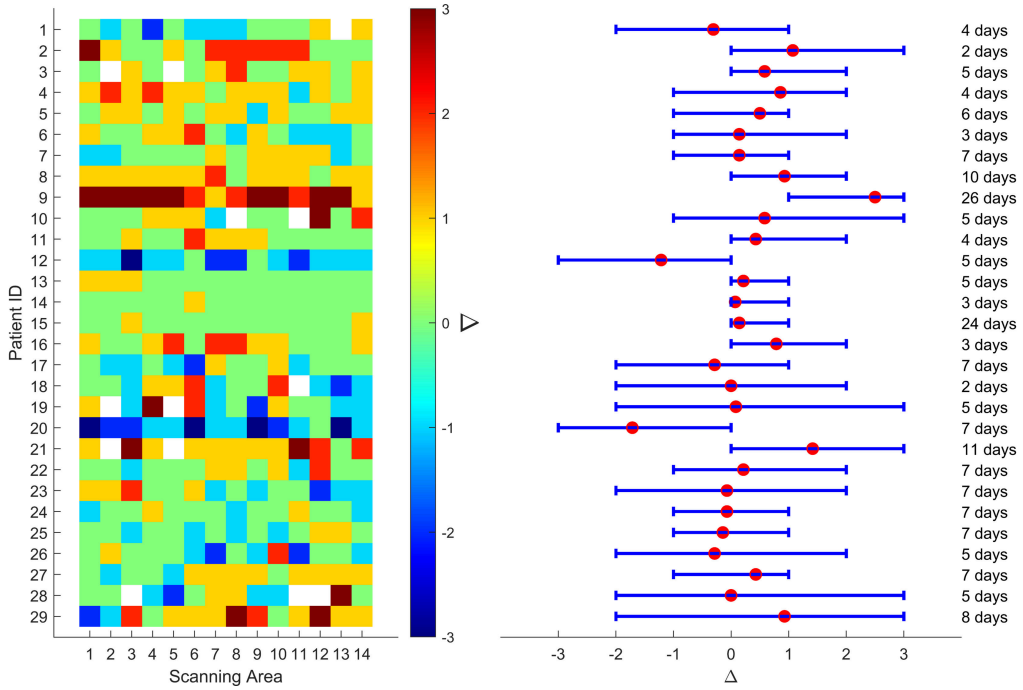

Figure 5.6: The values of Δ for each scanning area (x-axis) and for each patient (y-axis) that was scanned twice (on different dates) are depicted on the left. The 29 patients involved in this longitudinal study are numbered on the y-axis from 1 to 29. The white squares indicate the absence of the measurement. On the right side, the mean value of Δ for each patient is depicted with a red point, whereas the lower and upper bounds of each error bar represent the minimum and maximum Δ of each patient, respectively. The temporal distance (days) between the two LUS exams is indicated on the right.

and 68, 82, and 97% for post-COVID-19 patients (Figure 5.5b, top right). Consistently with a previous study [28], for this type of analysis the level of agreement was computed by summing the number of patients sharing the same worst score from the reference protocol and dividing it by the total number of patients.

Moreover, also the distributions of scores in the posterior areas show a similar trend when comparing COVID-19 patients (Figure 5.5a, bottom left) and post-COVID-19 patients (Figure 5.5b, bottom left). This is translated in consistent levels of agreement between system 4 and the modified systems 4 (10 areas instead of 14) when looking at COVID-19 patients (Figure 5.5a, bottom right) and post-COVID-19 patients (Figure 5.5b, bottom right).

Specifically, the levels of agreement between system 4 and the modified systems 4 are 83, 92, and 97% (COVID-19 patients), and 85, 86, and 96% (post-COVID-19 patients), when the scanned posterior areas were the apical, middle, and basal, respectively.

5.3.2 Longitudinal Study

As shown in Figure 5.6 (left), Δ values are generally heterogeneously distributed within each patient. As a consequence, the error bars are generally long (55.17% of error bars have a length equal to or greater than 3), with the mean values focalized in the center

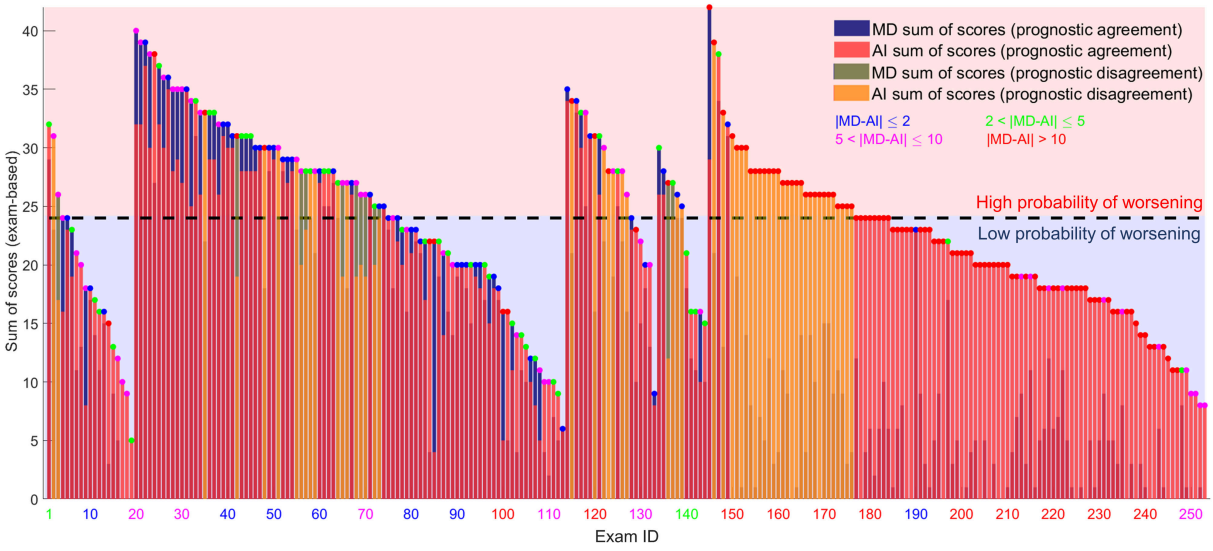

Figure 5.7: The exam-based sum of scores for each LUS exam are depicted. MD exam-based scores and AI exam-based scores are depicted in blue and red bars, respectively. Each exam is colored (colored points above each bar) in blue, green, purple, and red, depending on the disagreement interval. The bars highlighted in yellow represent the LUS exams where the prognostic evaluation of MD and AI differs. The dark dashed line indicates a cumulative score of 24, which defines the prognostic threshold. The five subgroups of exams have been divided as follows: COVID-19 patients from Rome (exam ID 1–19), Pavia (exam ID 20–113), and Lodi (exam ID 114–133), post-COVID-19 patients from Rome (exam ID 134–144), and Pavia (exam ID 145–253)

of each bar (Figure 5.6, right). It is noticeable that most of the mean values of Δ are positive (65.52%), meaning that most of the patients were recovering from the disease.

5.3.3 Prognostic Evaluation

Figure 5.7 shows how the prognostic agreement between AI and clinicians is higher in COVID-19 patients than in post-COVID-19 patients. Specifically, the prognostic agreement is 80.45% for COVID-19 patients (exam ID from 1 to 133) and 72.50% for post-COVID-19 patients (exam ID from 134 to 253). It is important to highlight how the AI models were trained on LUS data from COVID-19 positive patients.

As introduced in Section 5.2, the prognostic agreement was calculated assessing the capability of recently developed AI algorithms [2] of automatically stratifying patients at high risk of clinical worsening from patients at low risk. We hence compared, for each LUS exam, the cumulative score values (the sum of the scores over the 14-areas investigated) obtained from the analysis of the data performed by LUS experts, with that provided by the AI [45].

For post-COVID-19 patients, it is plausible to assume that there may be a moment in time in which the recovery process of the damaged lung tissue produces LUS patterns which are not fully compatible with those obtained from healthy or acute patients. This hypothesis could explain the different performance of the AI models on post-COVID-19 patients.

5.4 Discussion

In this study, we report on new results related to the application of a standardized LUS imaging protocol and scoring system, which was developed to assess LUS data from patients affected by COVID-19. The objectives of this study are multiple. First, to determine whether a simplified LUS imaging protocol could accurately capture and characterize the sonographic appearance of pleural and sub-pleural alterations in COVID-19 and post-COVID-19 patients. Standardization and evidence-based results are in fact fundamental with LUS, since one of the most important limitations of this type of exam is that it relies on qualitative and subjective interpretations of LUS videos, which are scored depending on the presence of relevant imaging patterns. Although standardization cannot remove subjectivity completely, it can help reducing it by defining how many areas need to be scanned and where, as well as detailing the range of key imaging parameters which should be utilized to acquire the data.

Beyond the extended dataset (220 patients), one of the novelties of this study are the comparison between data from COVID-19 and post-COVID-19 patients. From Figure 5.3, it is clear how post-COVID-19 patients present lower scores compared to COVID-19 patients. This is in line with the expectations, given the prevalence of nonhospitalized patients in the post-COVID-19 group.

From Figure 5.4, it is also interesting to note how the score distribution for COVID-19 patients was very similar among the different centers. Moreover, a clear difference could be observed between hospitalized and nonhospitalized post-COVID-19 patients, with the

latter subgroup displaying lower scores.

From Figure 5.5, we can observe how, for both COVID-19 and post-COVID-19 patients, the optimal trade-off in terms of amount of scanning areas is 12. Specifically, a level of agreement of 98 and 97% was respectively found when comparing the results with a 14-areas scanning protocol. Moreover, for both patients' populations, the worst scores are found in the basal posterior areas.

The second objective is to study whether the acquisition protocol could be further simplified during follow-up.

From Figure 5.6, although these results were obtained on a limited number of patients (29), it seems that it is not possible to derive the general evolution of the lung condition from a subset of areas. In fact, no strong correlation is found among values of Δ obtained over different areas. This implies that a 12-areas acquisition protocol is recommended also during follow-up.

The third objective is to investigate the performance of recently developed AI models to automatically assess LUS videos according to the above-introduced scoring system. AI models, especially when equipped with explainability mechanisms (which guarantee the possibility for a user to understand the decision made by the AI), can in fact further reduce the subjectivity of the evaluation process by providing a baseline evaluation. Moreover, automatic scoring algorithms can execute the analysis in a shorter time and relentlessly. They can thus be of great support for clinicians in case significant amount of data need to be analyzed in a short period of time.

From Figure 5.7, it is possible to showcase the potential as well as the limitations of AI models when applied to the analysis of LUS data. In fact, whether a good level of agreement was obtained between the AI and LUS experts in the evaluation of data from COVID-19 patients, the agreement was significantly reduced for post-COVID-19 patients. This is consistent with the fact that the employed AI models were trained only on data from COVID-19 patients. Once again, these results show how important it is to limit the application of AI models to the analysis of data consistently with the characteristic of the training dataset.

When presenting LUS findings, it is important to remember the current limitations of ultrasound technology when applied to the lungs. First of all, its low specificity. LUS patterns, as those investigated in this manuscript for COVID-19, are in fact nonspecific. It is thus fundamental not to misinterpret LUS as a tool applicable to diagnose COVID-19. Differently, it is applicable to evaluate the state of the lung and follow its evolution over time. It should also be acknowledged that ultrasound technology can only assess the surface of the lung, as the presence of air inhibits the exploration of deeper regions, unless the loss of aeration is significantly extended and reaches the surface of the lung. Another limitation of LUS, as performed through the analysis of data acquired with clinical scanners, is its intrinsic qualitative nature. In fact, although a numerical scoring system can be developed which associates specific patterns to a number, these approaches cannot be considered truly quantitative. To do that, measurable physical quantities with the power to characterize the alterations along the lung surface should be identified, and dedicated ultrasound methods designed around the peculiar properties of lung tissue should be developed. Research in this direction is emerging [58, 59], but further and extensive clinical

studies are necessary to define and validate the potential of these methodologies with respect to their reproducibility, accuracy, and specificity.

5.5 Conclusion

In conclusion, the proposed scoring system is applicable to assess COVID-19 and post-COVID-19 patients. For both COVID-19 and post-COVID-19 patients, a 12-areas acquisition protocol is confirmed as the optimal trade-off between a time-efficient and accurate LUS examination procedure. Moreover, the worst scores are confirmed to be found in the basal posterior areas for both patients' populations. As for what concerns follow-up studies, it appears not to be possible to simplify the acquisition process, as no clear correlation was found among the score evolution across different areas. Finally, LUS data obtained from COVID-19 and post-COVID-19 seem to display differences which are capable of confusing AI models that were not trained on post-COVID-19 data. This opens interesting questions on the existence of specific patterns associated to post-COVID-19 patients. Research in this direction will be the focus of future studies.

Chapter 6

On the influence of imaging parameters on lung ultrasound B-line artifacts, *in vitro* study

This Chapter[5] *presents an experimental quantitative study aiming at investigating the genesis of LUS vertical artifacts. Specifically, this study investigated an ambiguity about the formation of vertical artifacts, leading to the formulation of two main hypotheses. The first hypothesis states that the visualization of these artifacts is linked only to the dimension of the emitted beam, whereas the second associates their appearance to specific resonance phenomena. To verify these hypotheses, the frequency spectrum of these artifacts was studied by using dedicated lung-phantoms. A research programmable platform connected to an LA533 linear array probe was exploited both to implement a multi-frequency approach and to acquire raw RF data. The strength of each artifact was measured as a function of frequency, focal point, and transmitting aperture by means of the artifact total intensity. The results showed that the main parameter that influences the visualization of vertical artifacts is the frequency rather than the focal point or the number of transmitting elements. Moreover, the visualization of vertical artifacts seemed to be associated to specific resonance phenomena rather than the dimension of the emitted beam.*

6.1 Introduction

Thanks to their capacity to penetrate structures with similar acoustic impedances, such as soft-tissue, ultrasound waves are nowadays widely used in medical imaging in order to extract anatomical information of internal body parts. However, this is not possible when inspecting the lung. In fact, the acoustic impedance of the lung significantly differs from that of the soft tissues due to the presence of air and, hence, standard ultrasound imaging techniques are unable to directly provide anatomical images of this organ. In spite of this, ultrasound can provide useful information on the surface of the lung. This

[5]This Chapter appears in:
[J4] F. Mento, L. Demi, "On the influence of imaging parameters on lung ultrasound B-line artifacts, *in vitro* study," in *The Journal of Acoustical Society of America*, vol. 148, no. 2, pp. 975-983, August 2020.

information could be extracted by exploiting lung ultrasonography (LUS), which is based on the interpretation of the visual artifacts that appear in the reconstructed image, such as A-lines (also known as horizontal artifacts) and B-lines (also called vertical artifacts).

A-lines represent equally spaced horizontal repetitions of the pleural line, and their physical origin has been explained by the presence of two strong reflectors, i.e., the probe and the pleural line, which causes ultrasound waves to bounce between these two interfaces [24]. While the A-line pattern is mainly visible in healthy lungs, where the reflection coefficient of the pleural line can be approximated to 1, B-lines correlate with several pathological conditions of the lung [17, 21, 50, 108, 122]. Indeed, in these pathological conditions, part of the lung volume originally occupied by air is replaced by liquid or tissue, the original mismatch between the lung surface and the intercostal tissue layers is, thus, mitigated, and the organ becomes penetrable by ultrasound. This permeability can be associated with the formation of acoustic channels filled with media (tissue, water, or blood) where ultrasound can propagate [17, 24, 50, 87].

Even though several studies have investigated the origin of these artifacts by reproducing them in foamy phantoms [85], bubbly media [16, 55], and with rigid structures surrounded by a homogeneous medium [101], their formation remains unclear [23]. As a consequence, being limited to the visual evaluation of these artifacts leads to qualitative and subjective diagnoses, mainly based on the counting of vertical artifacts in the reconstructed image [123–125].

To quantitatively characterize B-lines, researchers have started to develop ultrasound imaging techniques specifically applicable to the lung [55–57]. However, further investigations are needed for a better comprehension of these artifacts.

This paper investigates the dependence of B-lines on different transmission parameters, i.e., frequency, focal point, and number of transmitting elements, in order to verify two main hypotheses, which arise from an ambiguity about the formation of these vertical artifacts. Since the dependence of B-lines on frequency has recently been discovered [55], the transmitted frequency of the pulse was considered as the main parameter in our analyses.

The first hypothesis states that the appearance of these artifacts is associated only to the size of the emitted beam; indeed, it assumes that B-lines are always visualized in the reconstructed image if the emitted beam is narrow enough to enter an acoustic channel with a sufficient quantity of energy. The truthfulness of this hypothesis would imply the possibility to adopt an approach based on thresholds, where the main target would consist in finding the lowest transmission frequency (and, thus, the larger beam size, assuming a fixed focal point and active aperture) at which the artifact becomes visible. This threshold will provide an indirect estimation of the acoustic channel size. Therefore, the appearance of B-lines at low frequencies (hence, for relatively large wavelengths) will imply a larger channel size and, therefore, signal a worse condition of the lung surface. A recently published review paper describes in details the importance of characterizing the acoustic channels in order to obtain diagnostic information on the status of the lung surface [54]. In contrast, the second hypothesis states that the presence of B-lines in the reconstructed image is due to specific resonance phenomena. In this case, the artifact should be clearly visible only around its resonance frequency with fading strength when

imaging with lower or higher frequencies. The parameter of interest for the geometrical characterization of the acoustic channel would then be the resonance frequency connected to each B-line artifact.

In this study, we investigate the validity of these two hypotheses by analyzing the frequency spectrum of B-lines observed in two lung-mimicking phantoms, which are able to reliably generate vertical artifacts [55]. The analyses were performed by means of the multi-frequency approach presented by Demi *et al.* [55], which we extended to higher frequencies and implemented on the ULA-OP research programmable platform [126]. The sum of intensities (referred as *"total intensity"*) is investigated in this paper to measure the artifact strength and exploited to analyze the B-line frequency trend (see Section 6.2.3). This feature does not capture the shape of the artifact and is not intended as an absolute measure of B-lines intensity but as a way to quantitatively compare the appearance of a B-line artifact with respect to the imaging frequency.

The paper is structured as follows. In Section 6.2.1, both the fabrication and characteristics of the phantoms are described, whereas Section 6.2.2 focuses on the description of the main equipment and the data acquisition procedure. Then, the process applied to compute the total intensity parameter is introduced in Section 6.2.3. Subsequently, the results are presented in Section 6.3, and the discussion and conclusion appear in Section 6.4.

6.2 Methods

6.2.1 Lung-mimicking phantoms

Three lung-mimicking phantoms were fabricated by trapping a population of monodisperse microbubbles, generated by the MicroSphere Creator (Tide Microfluidics, Enschede, the Netherlands), within a tissue-mimicking gel as done by Demi *et al.* [55]. The procedure used to fabricate the lung-mimicking phantoms can be resumed in five main steps. First, a rigid cylindric structure is partially inserted within the test tube, which has been previously filled up with the not yet solidified tissue-mimicking gel (first step). Subsequently, after the gel solidification, the cylinder is removed and, as a consequence, an empty channel is formed within the gelatin (second step). Then, this cavity is filled up with microbubbles by means of a hose (third step), and a relatively small layer of not yet solidified gelatin is added within the channel to partially trap these microbubbles (fourth step). Consequently, until the gelatin layer is not completely solidified, the microbubbles slowly migrate toward the gel surface and remain trapped when the solidification process is complete. This process generates a microbubble layer whose size along the z axis (depth) is about 5 mm for each phantom. However, the process does not allow each microbubble to be positioned in a precise location. Finally, after solidification, a larger quantity of tissue-mimicking gel is poured within the test tube to fill it (fifth step).

The employed microbubbles consisted of air-bubbles encapsulated in a lipid shell. To simulate the alveolar size reduction, which is typically observed in several pathological conditions [55], the diameter of these microbubbles was fixed to 80 μm, i.e., a smaller size compared to the normal alveolus, whose diameter is about 280 μm [11]. The microbubbles were embedded in tissue-mimicking gel with a speed of sound approximately equal to

that of the soft tissues (1540 m/s) [127]. However, the utilized tissue-mimicking gel does not provide attenuation values typical of soft tissues [127]. In fact, the tissue-mimicking gel's attenuation does not significantly vary with frequency (equals about 0.021 dB $cm^{-1}MHz^{-1}$), which is different from that of the soft tissues (approximately equal to 0.63 dB $cm^{-1}MHz^{-1}$) [127]. It is important to note that, given the diameter of the utilized microbubbles (80 μm), their resonance frequency will fall below 100 kHz (Ref. [101]) and, thus, far from the employed imaging frequencies. Figure 6.1 illustrates how the lung-mimicking phantoms are designed.

6.2.2 Data acquisition

The ULA-OP programmable platform [126] and an LA533 (Esaote, Florence, Italy) linear array probe, whose pitch and element size along the lateral dimension are, respectively, equal to 245 and 220 μm [128], were exploited to acquire data by means of a multi-frequency approach based on orthogonal sub-bands centered at different center frequencies. The utilized probe has a -6 dB bandwidth from 3.8 to 12 MHz and a -12 dB bandwidth from 3.2 to 13.2 MHz. The maximum of the transducer transfer function is at 8 MHz. To precisely evaluate the frequency spectrum of B-lines, we guaranteed each sub-band to have a narrow bandwidth (about 1 MHz at -10 dB) in the frequency domain by transmitting, for every center frequency, a Gaussian pulse having a 2 μs timelength. The most employed center frequencies were 3, 4, 5, and 6 MHz; however, to expand the spectral analysis presented by Demi *et al.* [55], four acquisitions were performed also at 8, 10, and 12 MHz (see Table 6.1). A 50 MHz sampling frequency was used. In reception, a sub-aperture of 64 elements was employed for all acquisitions. Each final image,

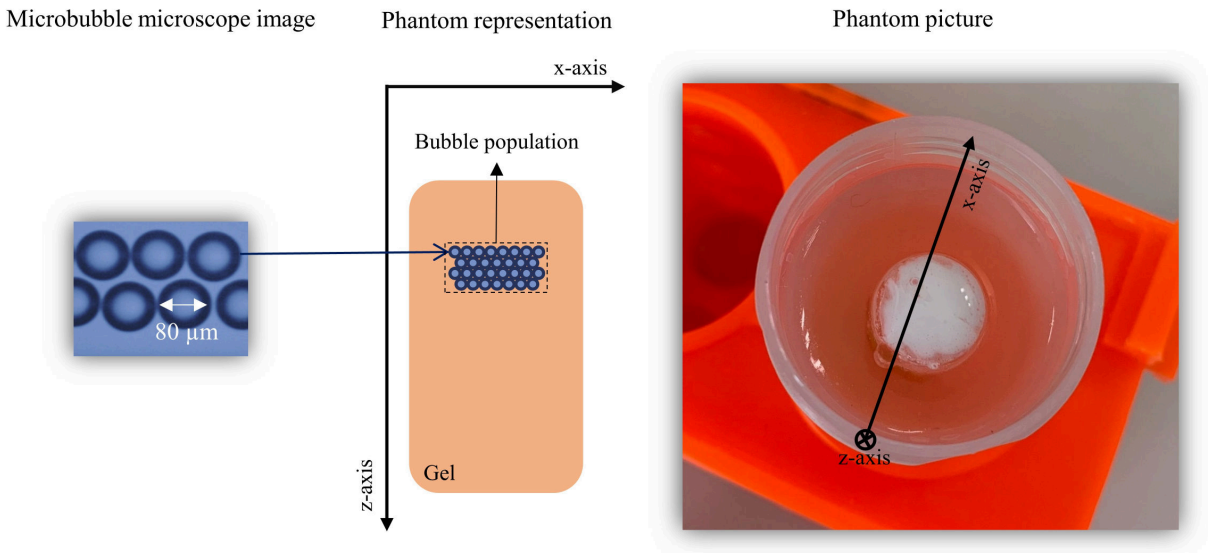

Figure 6.1: Microbubble microscope image, a phantom representation, and a top view picture of a lung-mimicking phantom. The x axis (lateral dimension) and the z axis (depth) are also represented.

which was reconstructed by linearly shifting this sub-aperture over the array, consists of 129 lines. The time gain compensation (TGC) was implemented as a linear function (in logarithmic scale) that increases as the depth increases. Since the attenuation does not significantly change with frequency in the tissue-mimicking gel surrounding the microbubble clouds [127], the TGC was kept unaltered for all the acquisitions. Thanks to an automatic positioning system allowing the management of the probe displacement, the phantoms, contained in a test tube rack and immersed in water, were visually inspected along their corresponding bubble layers to detect the presence of B-lines (see Fig. 6.2). Once a vertical artifact was detected, the probe was stopped, and the transmission parameters were varied as listed in Table 6.1. The number of transmitting elements was varied from 8 to 64, and the effect of the focal point location was evaluated for B-line IV. To guarantee the analyses to be robust, only particularly strong B-lines were considered. Therefore, the phantom having weaker B-lines was discarded, and the dataset was composed by considering the five artifacts observed in the other phantoms (see Table 6.1). Figure 6.3 shows an example of a reconstructed image, highlighting the main structures.

Furthermore, to objectively evaluate the first hypothesis, we measured the -12 dB lateral resolution by means of a wire phantom. To this end, a -12 dB lateral width of a wire target located at the same depth of the phantom 2's microbubble layer (i.e., 20 mm) and imaged by setting the same parameters utilized in the B-line III investigation

Number	Phantom	Elements	Focus (mm)	Frequency range (MHz)
I	1	8	20	3–6
I	1	16	20	3–12
I	1	32	20	3–12
I	1	64	20	3–12
II	1	8	20	3–6
II	1	16	20	3–6
II	1	32	20	3–6
II	1	64	20	3–6
III	2	64	22.5	3–12
IV	2	64	15	3–6
IV	2	64	17.5	3–6
IV	2	64	20	3–6
IV	2	64	22.5	3–6
IV	2	64	25	3–6
V	2	8	1000	3–6
V	2	16	1000	3–6
V	2	32	1000	3–6
V	2	64	1000	3–6

Table 6.1: The main parameters (number of transmitting elements, focus, and frequency) set in the different acquisitions are listed from the third to the last column. The first column indicates the identifying number of the analyzed B-line, whereas the second column indicates the analyzed phantom.

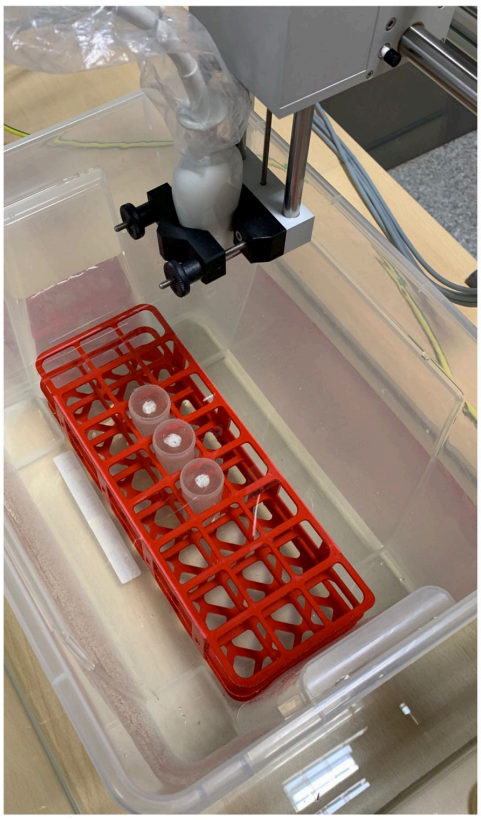

Figure 6.2: Picture of the acquisition system. The probe is controlled by means of the positioning system, and the lung-mimicking phantoms, which are contained within the test tube rack, are immersed in water.

(see Table 6.1) was measured. This analysis allows the comparison between the beam size at the microbubble layer's location and the total intensity as a function of frequency for B-line III (observed in phantom 2).

6.2.3 Computation of the total intensity parameter to quantify the strength of B-lines

First, for each artifact, 23 frames were acquired and averaged to obtain a final image for each center frequency. Then, to quantitatively evaluate the strength of the main B-lines (labeled from I to V in Table 6.1), we applied a four-step procedure to the final (averaged) images, as depicted in Fig. 6.4. Here, with *"main B-line"* we refer to the stronger and most clearly definable B-line present in an image. In step 1, both the bubble layer and the B-lines were included in the region of interest (ROI) with a rectangular box. In step 2, to exclude undesired frequency components, a sixth-order Butterworth bandpass filter (1.8 MHz bandwidth) centered at the transmitted center frequency was applied to the ROI, and the envelope was extracted by means of the Hilbert transform. Furthermore, the ROI was normalized with respect to its maximum value and visualized in logarithmic scale with a 35-dB dynamic range. In step 3, the main B-line was visually detected and

manually included in a smaller ROI (ROI_B) in order to isolate it from both the secondary B-lines and the bubble layer. Finally, the strength of the main B-line was quantitatively evaluated by introducing the total intensity parameter [59], which was computed in step 4 as follows. As a preliminary operation, we set an arbitrary threshold at -35 dB to consider only the values above the threshold in the computation of the total intensity. Subsequently, we calculated the total intensity in logarithmic scale as

$$I_{TOT} = 20\log_{10}(A_{pix}\sum_{i,j} 10^{\frac{ROI_B(i,j)}{20}}) \tag{6.1}$$

where A_{pix} is the area of a pixel (equal to 3.6×10^{-3} mm^2), i and j indexes refer to

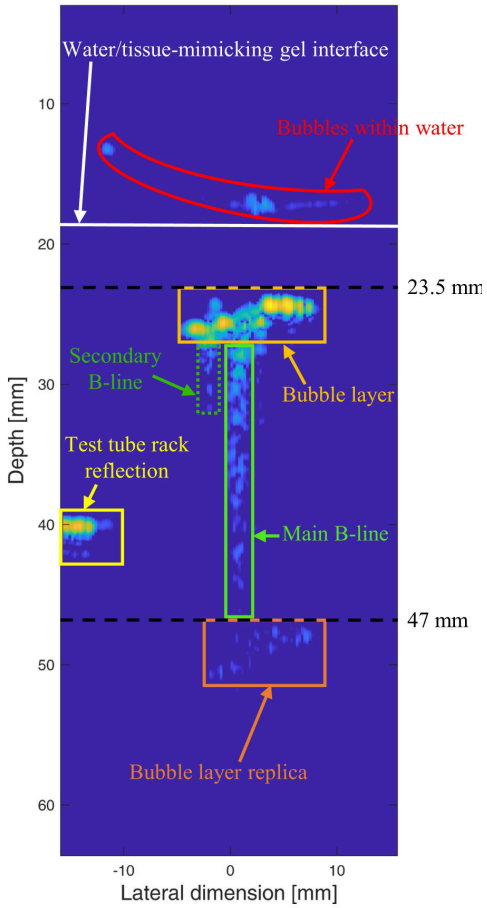

Figure 6.3: The main elements in the reconstructed image are shown. The white line represents the estimated depth of the water/gelatin interface, whereas the red shape highlights the location of bubbles formed within water. The first strong reflection, associated to the bubble layer, appears at 23.5 mm, and its replica (A-line) appears at 47 mm. The green boxes contain the main B-line (solid-line box), i.e., the analyzed artifact, and a secondary B-line (dashed-line box). The reflection caused by the test tube rack is highlighted by the yellow box. The image is shown with a 35-dB dynamic range.

the pixel in ROI_B located at the ith row and jth column, and $ROI_B(i,j)$ is the intensity value (in dB) of the pixel located at the ith row and jth column. Given a consecutive block of N images obtained at the N different frequencies (where $N = 4$ when the used center frequencies are only 3, 4, 5, and 6 MHz, and $N = 7$ when imaging also at 8, 10, and 12 MHz), the ROI was manually selected from the image with the largely extended B-line, starting below the microbubble layer and in order to contain the entire artifact as visualized from a normalized image with a dynamic range of 35 dB. The same ROI was then utilized for the other N-1 images of the block.

The total intensity parameter introduced by Eq. 6.1 represents, therefore, a metric able to quantify the B-line strength, implicitly considering both the artifact spatial extension and intensity. Moreover, the normalization applied in step 2 ensures the independence of this parameter from the received signal amplitude, which can be lower at higher frequencies due to, e.g., the probe transfer function. This increases both the reliability and the robustness of the analysis. It is worth mentioning that although the pixel area does not vary within the data presented in the paper, this parameter is generally different for different imaging systems.

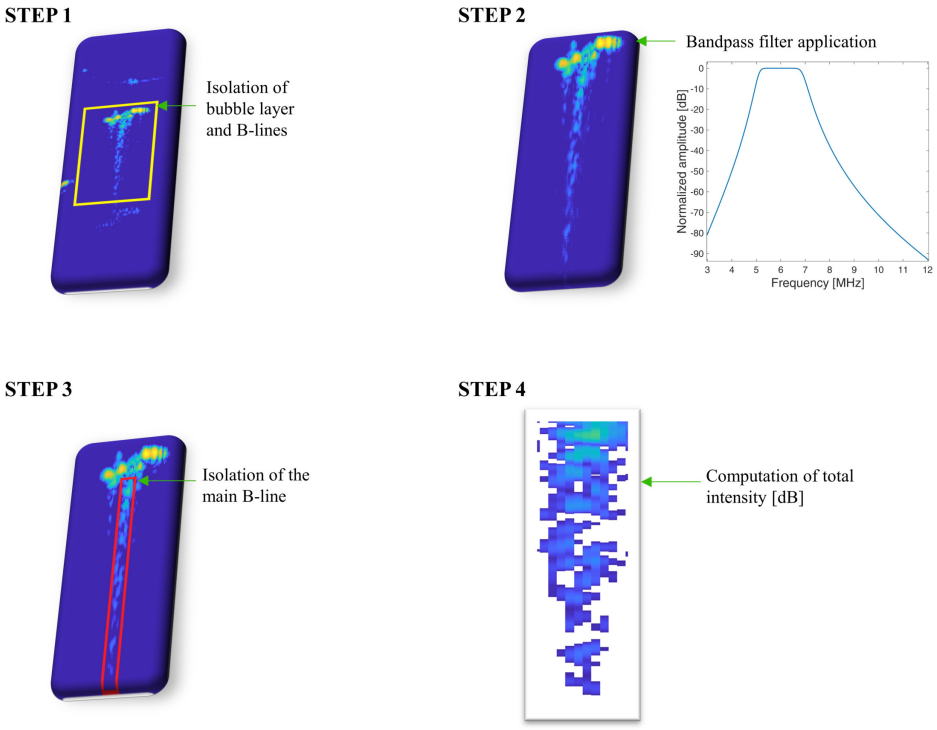

Figure 6.4: Schema of the procedure applied to the final (averaged) images to compute the total intensity of each main B-line. For simplicity, only the bandpass filter employed when imaging at 6 MHz is depicted in the figure. The filter is obviously adapted, depending on the specific imaging frequency. The images are shown with a 35-dB dynamic range.

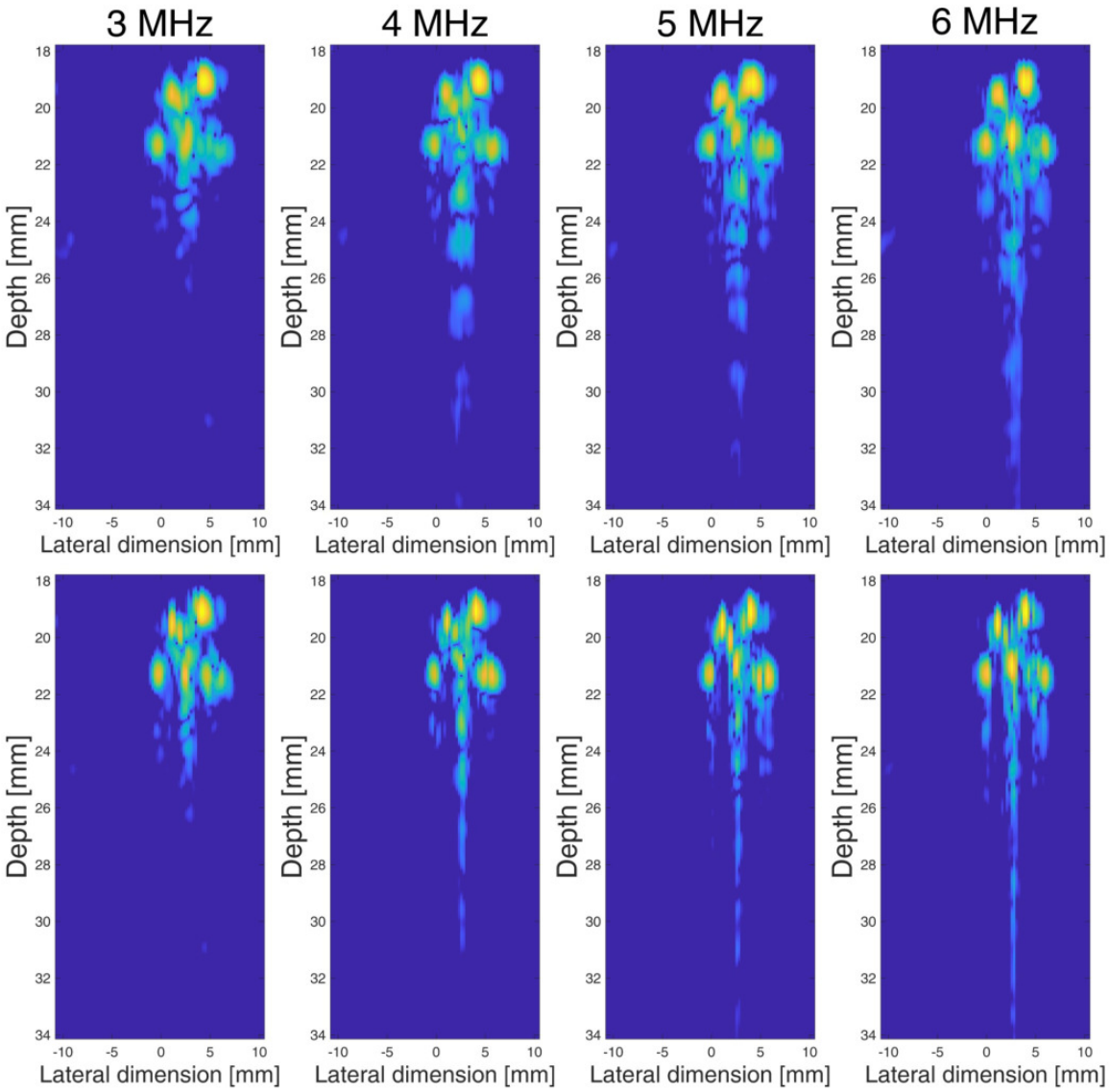

Figure 6.5: Multi-frequency images of phantom 2 (B-line IV) with the focus at 15 mm (top) and 17.5 mm (bottom). The images are shown with a 35-dB dynamic range.

6.3 Results

Figure 6.5 shows an example of multi-frequency images obtained with different focal points. It is observable how a 2.5 mm displacement of the focus qualitatively changes the images.

The frequency dependence of B-lines can be observed more in detail in Fig. 6.6, where multi-frequency (from 3 to 12 MHz) images of phantom 1 (top) and phantom 2 (bottom) are shown. Both phantoms show a specific spectral behavior.

To quantitatively evaluate the frequency spectra of B-lines, we considered the nor-

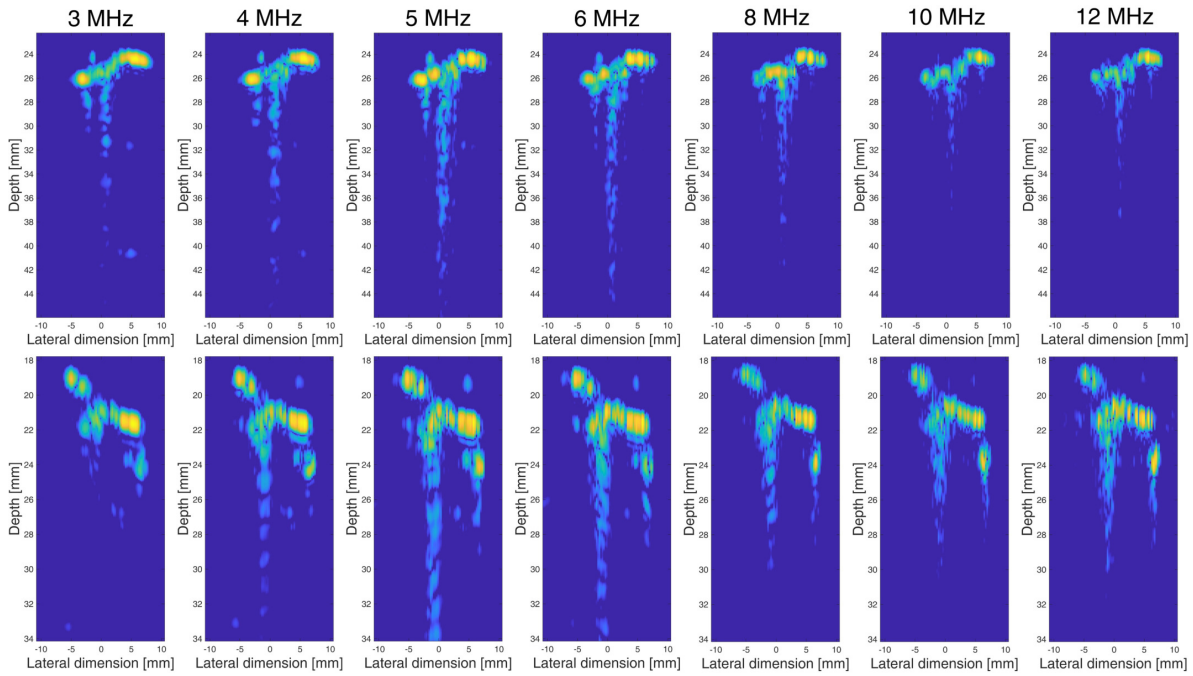

Figure 6.6: Multi-frequency images of phantom 1 (B-line I) and phantom 2 (B-line III) are shown at the top and bottom, respectively. Both acquisitions were performed with 64 transmitting elements. The images are shown with a 35-dB dynamic range.

malized total intensity, i.e., a scaled version of the total intensity parameter. We simply applied a normalization with respect to the maximum value of the total intensity measured by varying the transmission frequency and maintaining the other parameters constant.

Figure 6.7 shows the normalized total intensity of B-line I (phantom 1) as a function of frequency, also displaying the influence of the number of transmitting elements on the artifact. The mean trend of the normalized total intensity, depicted as blue points in the top graph, can be considered as bell shaped. Its maximum value (-0.55 dB) is at 6 MHz, whereas its minimum value (-10.83 dB) is at 12 MHz. As shown in Fig. 6.7 (bottom), the frequency trend of the normalized total intensity remains approximately the same, regardless of the number of transmitting elements (which affects both the beam width and incident pressure). However, considering adjacent frequencies (e.g., 3 and 4 MHz), a higher number of transmitting elements leads to a higher difference of the total intensity. This difference can be highlighted by comparing, for example, the results obtained with 32 and 64 elements at 3, 4, 5, and 6 MHz, where the normalized total intensity is, respectively, equal to -4.24, -2.90, -0.71, and 0 dB for 32 elements and equal to -8.34, -4.86, 0, and -2.21 dB for 64 elements. Nevertheless, the maximum variation of normalized total intensity caused by the different number of transmitting elements is significantly lower than the maximum variation caused by the frequency. Indeed, the former is approximately equal to 5.80 dB (error bar at 10 MHz in Fig. 6.7), whereas the latter is about 13.06 dB.

The frequency spectrum of phantom 2 (B-line III) is shown in Fig. 6.8. Also in this case, the frequency trend is substantially bell shaped, but visibly different from the trend

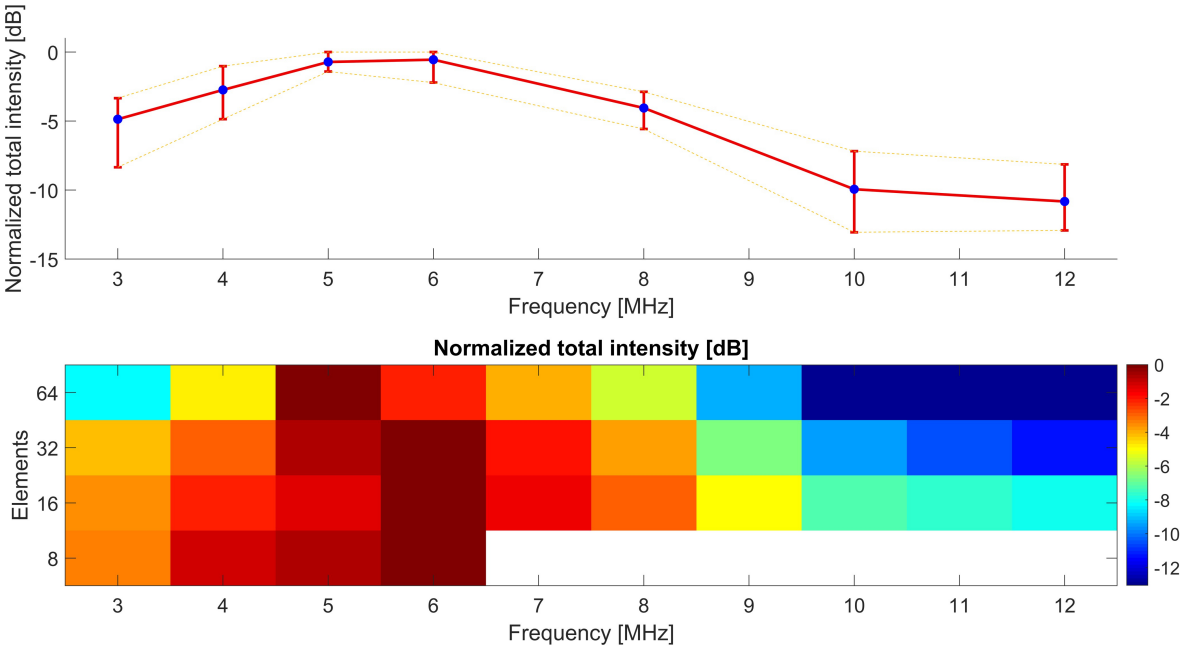

Figure 6.7: The normalized total intensity of phantom 1 (B-line I) as a function of frequency. The matrix (bottom) represents the normalized total intensity obtained from acquisitions performed with a different number of transmitting elements (rows). The white cells denote unperformed acquisitions, whereas the values at 7, 9, and 11 MHz are obtained with a linear interpolation. In the graph (top), the blue points represent the mean of each column of the matrix, whereas the error bars indicate the minimum and the maximum of each column (values obtained with linear interpolation are not explicitly represented in the graph).

of phantom 1 (Fig. 6.7). As expected, the visual imperceptibility of B-line III at 3 MHz (Fig. 6.6) is translated into a significantly low value of the normalized total intensity (-26.95 dB). On the other hand, the normalized total intensity at 4, 5, 6, 8, 10, and 12 MHz is, respectively, -5.72, 0, -1.61, -8.36, -9.45, and -6.51 dB. Therefore, contrary to phantom 1 (B-line I), the native frequency, which we consider as the frequency where the total intensity is maximum, is 5 MHz. Moreover, Fig. 6.8 shows the corresponding -12 dB lateral resolution at the microbubble layer's position.

Finally, we quantitatively compare the influence of the different transmission parameters (focal point, number of elements, and frequency) on the B-line total intensity. Results are shown by the boxplot in Fig. 6.9. The frequency box considers total intensity variations of B-line I and III (frequency range from 3 to 12 MHz), whereas the focus box includes variations of B-line IV (focal point from 15 to 25 mm), and the elements box considers variations of B-line I, II, and V (number of transmitting elements from 8 to 64). The total intensity variation is defined as the maximum variation of a given B-line total intensity with respect to a specific parameter. For example, the total intensity variation for B-line III with respect to frequency is 26.95 dB (see Fig. 6.8).

Hence, each box represents the maximum variations of the normalized total intensity

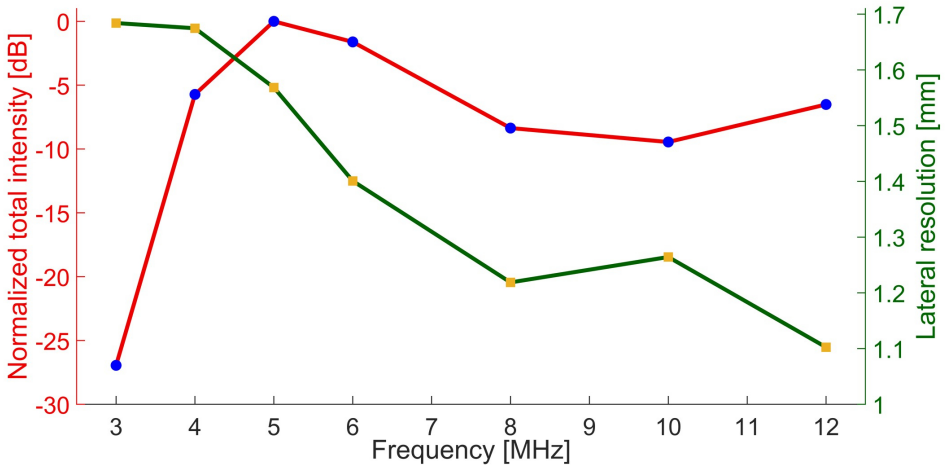

Figure 6.8: The normalized total intensity of phantom 2 (B-line III) as a function of frequency (red line with blue points) and the -12 dB lateral resolution measured by means of a wire phantom and by setting the same parameters of B-line III acquisition (green line with orange squares). The acquisitions were performed with 64 transmitting elements.

obtained by keeping fixed all the imaging parameters except from the investigated one. Figure 6.9 confirms that the frequency has a greater impact on B-lines compared to the other imaging parameters; indeed, the frequency box median value is equal to 11.42 dB, whereas the others do not exceed 4.5 dB. Moreover, the frequency box has both the widest interquartile range (9.59 dB) and the highest maximum value (26.95 dB).

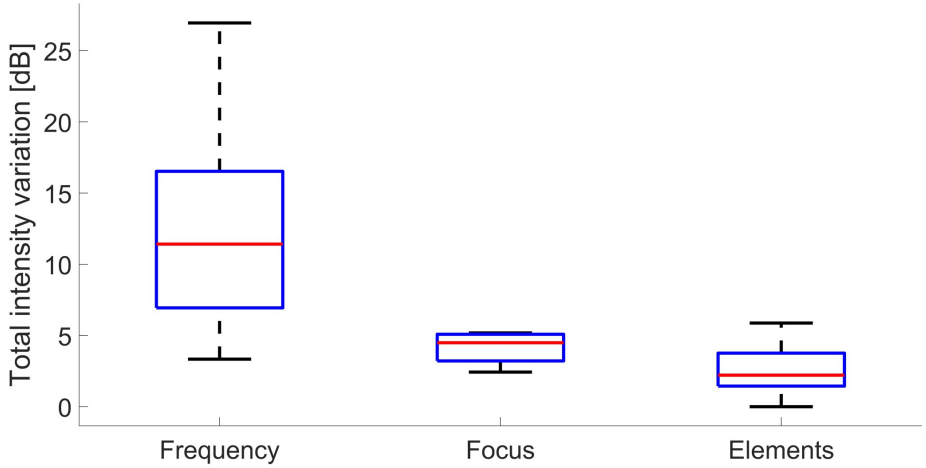

Figure 6.9: Boxplot showing the total intensity variations caused by frequency (left), focus (center), and number of transmitting elements (right). The red lines represent the median values, and the inferior and superior limits of the box are, respectively, the 25th and 75th percentiles, whereas the maxima and the minima are represented by horizontal black lines.

6.4 Conclusions and Discussion

Lung ultrasound imaging nowadays provides important clinical information on the state of the lung to physicians. However, due to the lack of understanding on the formation process of B-lines [23], the diagnoses are qualitative and subjective. Therefore, to base the clinical evaluation on a quantitative measure, a deeper comprehension of the mechanisms behind the generation of these vertical artifacts is needed.

In this paper, we have presented an *in vitro* study that investigates the influence of different transmission parameters on B-lines to verify two hypotheses about the formation of these artifacts.

Independently from the frequency, the first hypothesis links the B-line visualization only to the emitted beam size. In this case, the greater the energy (relatively to that associated to a given beam) entering an acoustic channel or trap, the more intense the B-line generated by the interaction between the transmitted wave and the channel will be. Hence, the acoustic channel aperture could be indirectly measured by controlling the beam size and, at the same time, disclosing the presence or absence of the vertical artifact in the image. If this hypothesis were true, increasing the transmit aperture size while keeping the imaging frequency constant or increasing the imaging frequency while keeping the aperture size constant should result in increasing the B-line total intensity as in both cases the beam width is generally reduced.

In contrast, the second hypothesis associates the B-line visualization to specific resonance phenomena, which could be exploited to characterize the acoustic properties of the channel rather than only its aperture size. If this hypothesis were true, the transmit aperture size should not significantly impact the B-line intensity, and the imaging frequency would be the most significant parameter for B-line characterization.

To investigate these two hypotheses, we have exploited the total intensity parameter to evaluate the frequency spectra of five B-lines detected in two different lung-mimicking phantoms. Thanks to the normalization with respect to the maximum value of each image, this parameter allows the evaluation of the B-line intensity independently from the absolute signal strength, hence, guaranteeing a reliable comparison between data acquired with different transmission parameters.

The results show a clear inconsistency between the frequency trend of B-lines (Figs. 7 and 8) and the beam size, which generally decreases as the transmitted frequency increases, as also shown in Fig. 6.8. As a consequence, the first hypothesis cannot always be considered to be true. Indeed, if this hypothesis were valid, we should expect a growing total intensity with increasing frequency due to the decrease of the beam size. On the contrary, we have shown the existence of cases where B-lines can be clearly visualized only within a specific frequency range (Fig. 6.6). Accordingly to these experimental results, the frequency response of these bubbly structures is more likely associated to specific resonance phenomena (second hypothesis).

Moreover, it is important to note that the highest total intensity is always measured at 6 or 5 MHz (Figs. 7 and 8), hence, confirming the same native frequencies observed before for bubbly structures generated with this size of microbubbles, although they had been measured with a different probe [55]. This further strengthens the hypothesis that the B-line frequency spectrum can be utilized to quantitatively characterize bubbly structures.

In addition, the results (Fig. 6.9) show the importance of exploiting the transmitted frequency, rather than the number of transmitting elements and the focal point, to extract important information from B-line artifacts. The variance of the B-line total intensity is indeed extremely larger when the frequency is considered as the main cause of variation. For this reason, the transmitted frequency appears to be the most relevant parameter in both the visualization and characterization of B-lines.

In conclusion, the spectral analysis of B-line artifacts may, thus, play an important role for the estimation of the state of the lung parenchyma. In this paper, we have introduced a quantitative parameter that allows us to evaluate *in vitro* the frequency spectrum of different B-lines. Future work will focus on the investigation of the diagnostic information that can be extracted from the B-line frequency spectrum. As there exists a clear difference of B-line intensity between cardiogenic and pneumogenic interstitial syndromes [24], further studies could investigate the possibility to quantitatively differentiate between these two pathological conditions.

The lung-phantoms described in this paper represent a well-controlled yet simplified model, which tries to approximate the lung parenchyma. As a consequence, the findings described in this paper may not directly apply to clinical data. To this end, future work will focus on improving these studies by, for example, analyzing the impact of polydisperse versus monodisperse bubble populations, the effect of including tissue-mimicking structures modeling the intercostal tissue layers, and the importance of bubble properties, such as shell material and thickness. Moreover, similar studies on radio frequency (RF) clinical data are essential, and currently ongoing. However, it is important to emphasize how the development of tissue-mimicking phantoms is essential for studies like the one presented in this paper. The absence of motion is, in fact, essential to guarantee that the exact same portion of the volume of interest is evaluated under the different settings. Moreover, control on bubbles size is another important aspect to acknowledge.

Furthermore, numerical simulations could be of strong interest to further verify the findings of this paper and deepen our understanding of the physical phenomena involved in the genesis of B-lines [58]. Numerical experiments would indeed be useful to study the interaction between ultrasound waves and acoustic traps with specific sizes and shapes. Successively, the numerical results could be verified in experimental models by finding a strategy to control the microbubble spatial distribution, thus, generating precise and well-defined acoustic traps. Another interesting study to be performed consists of experimentally evaluating (with dedicated phantoms) the influence of the angle of incidence of ultrasound beams on the appearance and visualization of vertical artifacts.

Chapter 7

Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: An *in vitro* study

This Chapter[6] *presents experimental quantitative study investigating the dependence of LUS vertical artifacts on different imaging parameters. Indeed, to develop quantitative techniques is fundamental to understand which parameters influence the vertical artifacts' intensity. In this study, we quantitatively analyzed the dependence of nine vertical artifacts observed in a thorax phantom on four parameters, i.e., center frequency, focal point, bandwidth, and angle of incidence. The results showed how the vertical artifacts are significantly affected by these four parameters, and confirm that the center frequency is the most impactful parameter in artifacts' characterization. These parameters should hence be carefully considered when developing a LUS quantitative approach.*

7.1 Introduction

The use of lung ultrasound (LUS) in clinical practice is nowadays widely spread, also due to the coronavirus disease 2019 (COVID-19) pandemic [1, 32]. In fact, the characteristics of LUS (e.g., real-time imaging, portability, and wide availability) render it particularly useful in emergency contexts [32, 42].

Despite its current spread, LUS remains strongly qualitative and subjective, as it is mainly based on the visual interpretation of imaging artifacts. Of particular importance are vertical artifacts, which correlate with various pathologies [21, 50, 108] and whose genesis is still not fully uncovered [24, 42]. Currently, only the so-called semi-quantitative approaches can be used to reduce the impact of subjectivity on the LUS exam, through

[6] This Chapter appears in:
[J5] F. Mento, L. Demi, "Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: An *in vitro* study," in *The Journal of Acoustical Society of America*, vol. 150, no. 6, pp. 4075-4082, December 2021.

the standardization of the acquisition and evaluation processes [1, 42]. However, also semi-quantitative techniques present several points of improvement, as they do not directly rely on the estimation of physical quantities. They indeed do not exploit the main characteristics of vertical artifacts, such as their dependence on different imaging parameters [55, 60]. These approaches are in fact implemented on images generated with the currently available clinical scanners, which are not designed for investigating the lung [42].

To further reduce subjectivity, quantitative approaches [55–60] are essential, as they aim at estimating physical quantities able to characterize the signals backscattered from the lung, as those responsible for the visualization of LUS vertical artifacts [42]. However, the development of quantitative approaches requires first a deeper comprehension of vertical artifacts' genesis, which has been investigated *in vitro* [16, 55, 60, 85, 101], and recently also *in silico* [96]. The appearance of these artifacts seems to be linked to the formation, along the lung surface, of acoustic traps [24, 50, 54, 87], which through resonance phenomena [60] can give rise to vertical artifacts. Moreover, as shown in a previous study [60], the transmitted frequency seems to be the most impactful parameter to characterize this type of artifact. However, it is necessary to also analyze the impact of other parameters on the vertical artifacts' strength, as to study the stability of a frequency characterization as well as to support the standardization of LUS and the development of quantitative approaches.

In this *in vitro* study, we aim at investigating the dependence of vertical artifacts on four different parameters, i.e., center frequency (CF), focal point (FP), bandwidth (BW), and angle of incidence (AOI). Even though we have already studied the dependence of vertical artifacts on CF and FP [60], the hereby utilized phantoms are completely different, as here we exploited a commercial thorax phantom used to train clinicians and not phantoms made with controllable monodisperse microbubble populations [60]. Another element of novelty is the investigation of the vertical artifacts' dependence on BW and AOI (never reported before). The latter is particularly interesting, as it could be a source of significant artifacts' variability during a LUS examination, especially when a convex probe is used.

The paper is organized as follows. Both materials and data acquisition process are presented in Section 7.2.1, whereas the procedure used to quantify the strength of vertical artifacts is described in Section 7.2.2. Successively, the results are shown in Section 7.3, and the discussion and conclusions in Section 7.4.

7.2 Materials and Methods

7.2.1 Materials and data acquisition

To perform this study, we exploited a Blue Phantom COVID-19 Ultrasound Lung Simulator (CAE Healthcare, Sarasota, FL), which consists of a thorax phantom able to reproduce LUS vertical artifacts, and generally used to train clinicians to use LUS in patients affected by COVID-19 pneumonia. As shown in Fig. 7.1, the phantom was positioned vertically within a water tank, and partially immersed in water. A metallic plate was used to keep the phantom's position fixed. The linear probe's displacement was precisely managed by means of an automatic positioning system (GAMPT, Merseburg, Germany).

A rotation system (GAMPT, Merseburg, Germany) was exploited to change the angle of incidence of the ultrasound beam along the x–z plane. A total of eight different scanning areas (from Area 1 to Area 8), including nine different vertical artifacts, were investigated in this study. These areas were chosen by visually inspecting the phantom with the probe along different positions of the x–y plane. The scanning areas showing vertical artifacts were then selected.

Data were acquired by exploiting an ULA-OP programmable platform [126] and an LA533 (Esaote, Florence, Italy) linear array probe (pitch and element size along the lateral dimension are 245 and 220 μm [128]) whose –6 dB bandwidth is from 3.8 to 12

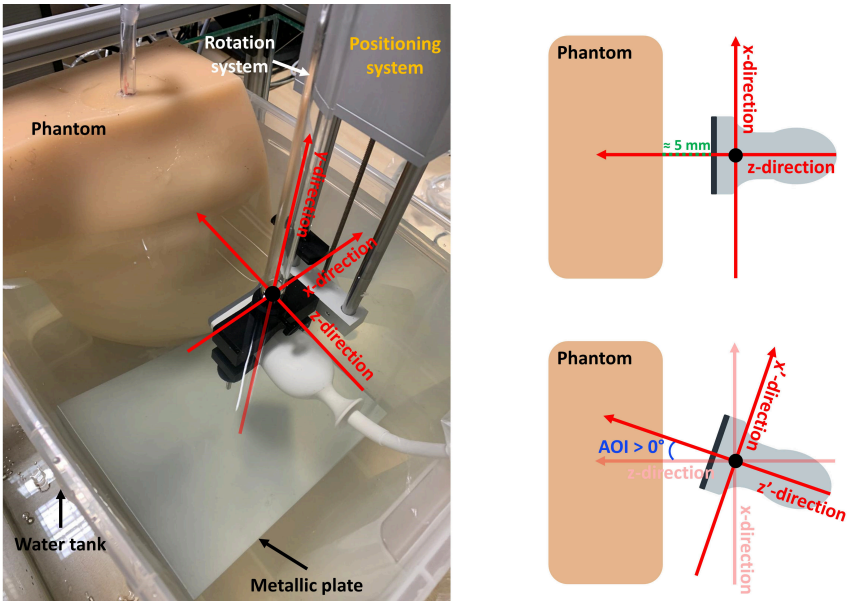

Figure 7.1: The utilized phantom, probe, positioning system, rotation system, metallic plate, water tank, and the three directions (x, y, and z) are shown on the left. A sketch showing (from a top-view) the geometry of the acquisition is depicted on the right. The distance between the phantom and the probe is approximately equal to 5 mm, when the AOI is 0°. The bottom right sketch shows how the acquisition geometry changes when AOI changes. Specifically, the x-direction and the z-direction are tilted by an angle equal to AOI (represented by the x'-direction and z'-direction in the bottom right sketch)

Analyzed parameter	CF [MHz]	BW [MHz]	FP [mm]	AOI [°]
CF	3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5	0	30	0
CF	3, 4, 5, 6	1	30	0
CF	4, 6	2	30	0
BW	4	0.5, 1, 2	30	0
FP	4	2	10, 20, 30, 40, 50	0
AOI	4	2	30	-30, -20, -10, 0, 10, 20, 30

Table 7.1: The parameters set for each scanning area. The first column indicates the investigated parameter (the parameter that was varied, keeping the remaining ones fixed), i.e., CF, BW, FP, and AOI. The values of each parameter are indicated from the second column to the fifth column.

MHz, and its –12 dB bandwidth from 3.2 to 13.2 MHz [60]. The transducer transfer function has its maximum at 8 MHz [60]. The radio frequency (RF) data were acquired by varying the parameters of interest as described in Table 7.1. For each scanning area, the center frequency was varied from 3 to 6.5 MHz, the bandwidth was evaluated at 0.5, 1, and 2 MHz, the focal point was set from 10 to 50 mm with 10-mm step size (the pleural line is approximately at 30 mm of depth), and the angle of incidence varied from –30° to 30° with a 10° step size. The choice of these center frequencies was made based on frequencies generally used in clinical practice, and by considering the probe transfer function. Furthermore, we chose CF to be consistent with previous clinical studies [58, 59]. The bandwidth was set by considering the probe transfer function, whereas the focal point was arbitrarily and significantly varied by considering the pleural line's depth as the main focal point (the structures to be imaged were at pleural line). The angle of incidence was varied so as to evaluate a significant change in the volume of interest's geometry and, at the same time, avoid the evaluated vertical artifact to be outside the volume of interest when changing AOI. A Gaussian pulse having a time length equal to 1, 2, or 4

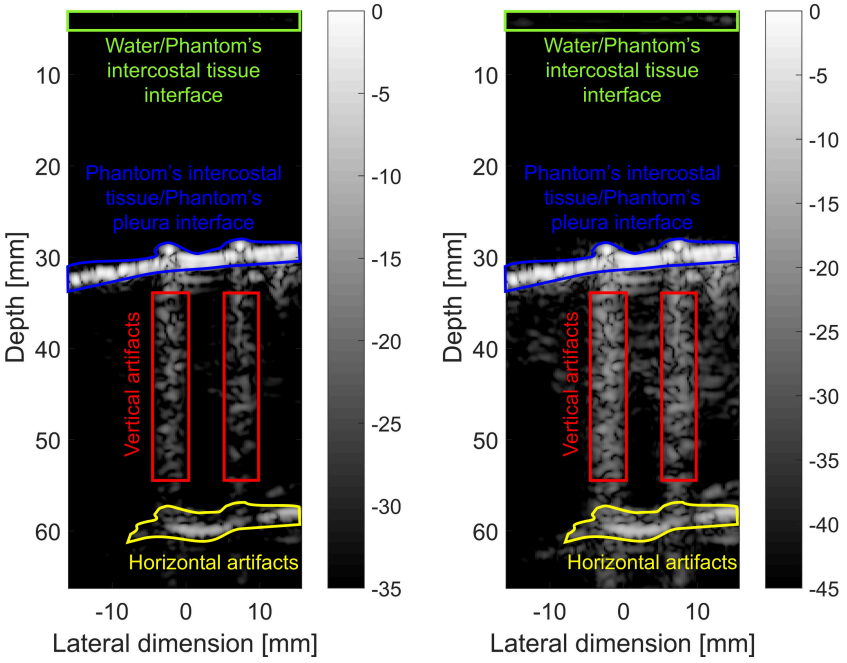

Figure 7.2: The main elements visualizable in the reconstructed image are shown. The shown image was acquired from the first scanning area (Area 1) with CF = 3 MHz, BW = 1 MHz, FP = 30 mm, and AOI = 0°. The Water/Phantom's intercostal tissue interface (in green) is at about 5-mm depth. The Phantom's intercostal tissue/Phantom's pleura interface (in blue) is at about 30-mm depth. The vertical artifacts (in red) are between approximately 35 and 55 mm of depth. The horizontal artifacts (in yellow) showing replicas of the Phantom's intercostal tissue/Phantom's pleura interface are about 60 mm of depth, i.e., double the distance between the Phantom's intercostal tissue/Phantom's pleura interface and the probe. The same image is shown at 35-dB dynamic range (left) and 45-dB dynamic range (right) to better visualize the Water/Phantom's intercostal tissue interface.

μs (depending on the desired bandwidth, i.e., 2, 1, or 0.5 MHz) was transmitted, and a 50 MHz sampling frequency was used. A sub-aperture of 64 elements was employed in reception. Hence, to reconstruct the final image, this sub-aperture was linearly shifted over the entire array (192 elements), thus obtaining an image with 129 lines along the lateral dimension [60]. For the time-space conversion (along depth), we assumed a speed of sound equal to 1546 m/s in the volume of interest. Figure 7.2 shows an example of reconstructed image with the main elements highlighted. From top (shallow) to bottom (deep), it is possible to observe the first acoustic interface dividing the water from the phantom's intercostal tissue (at about 5 mm), the second acoustic interface (i.e., the pleural line) dividing the phantom's intercostal tissue from the phantom's pleura (at about 30 mm), the vertical artifacts (between approximately 35 and 55 mm), and the horizontal artifacts showing replicas of the second interface (at about 60 mm, i.e., double the distance between the second interface and the probe).

To apply an adequate time gain compensation (TGC), we needed to estimate the attenuation coefficient [129] of the phantom. Therefore, we placed the phantom at a fixed position, and acquired RF data of the two acoustical interfaces, i.e., the water/phantom's intercostal tissue interface and the phantom's intercostal tissue/phantom's pleura interface, by setting the center frequency from 3 to 6.5 MHz (BW= 0.5 MHz, FP = 30 mm, AOI = 0°; see first acquisition setting in Table 7.1). The RF data were acquired avoiding saturation phenomena, and the difference of intensity between the first and second interfaces was computed for each center frequency. Given the distance between the two interfaces, we estimated the attenuation coefficient, which was equal to 2.15 dB/MHz cm. This value was hence adopted for the TGC.

To minimize variations of the pressure amplitudes delivered at the pleural line, we measured the received signal amplitude with changing frequency by means of a wire phantom. Specifically, we placed an 80-μm wire (immersed in water) at the pleural line depth (about 30 mm), and acquired RF data by setting CF from 3 to 6.5 MHz (0.5-MHz step), bandwidth equal to 0.5 MHz (see first acquisition setting in Table 7.1), and fixed the driving signal amplitude to 80% of the maximum amplitude allowed by the ULA-OP system. With driving signal, we intend here the electrical signal utilized to excite each element of the transmit aperture. Next, we adapted the transmitted amplitudes of the driving signals as to normalize the amplitude of the signals received at the different frequencies. This operation allows us to compensate for the probe transfer function, and reduce the effect of saturation phenomena.

7.2.2 Quantification of vertical artifacts' strength

As done in a previous study [60], the final image was obtained by averaging 23 frames, for each scanning area and each parameter setting. Then, a three-step procedure [59] was applied to the final (averaged) images to extract the total intensity (I_{TOT}) parameter, which allows us to quantitatively evaluate the vertical artifacts' strength [59, 60]. In step 1, a sixth order bandpass Butterworth filter centered at the different center frequency was applied [59, 60]. Specifically, the MATLAB *butter* function was utilized. The filter's bandwidth was adapted to the bandwidth of the signal, i.e., a filter's bandwidth equal to 1, 1.8, and 3.2 MHz was used when the transmitted signal's bandwidth was 0.5, 1, and

2 MHz, respectively. We successively applied the Hilbert transform to the filtered image, thus extracting the envelope, and normalized the image with respect to its maximum value [59, 60]. In step 2, the images were displayed in logarithmic scale with a 35-dB dynamic range, to visually detect the clearest vertical artifact [59, 60]. Each vertical artifact was segmented by selecting a rectangular region of interest (ROI_B) [59, 60], which was defined from the following acquisition setting: CF = 3 MHz, BW = 0.5 MHz, AOI = 0°, FP = 30 mm, i.e., where the vertical artifact was more spatially extended. For the same vertical artifact, the ROI_B was maintained fixed for each acquisition setting [59, 60], except from the acquisitions performed with AOI≠0°, as the geometry changed. In this case, the position of ROI_B was manually adapted to entirely include the vertical artifact as visualized at a 35-dB dynamic range. In the final step, where I_{TOT} is computed, the empirical threshold definition is fundamental (the empirical threshold defines the minimum value that has to be considered in the computation of I_{TOT} [59, 60]). As the noise level was empirically estimated to be at -45 dB, the arbitrary threshold was differently set from -33 to -37 dB (1-dB step size), obtaining five different values of I_{TOT} for each artifact and each parameter setting. The intention here is to analyze the impact of the threshold on the results.

7.3 Results

To quantitatively evaluate the dependence of the vertical artifacts' I_{TOT} on the four investigated parameters (CF, AOI, BW, and FP), we exploited a scaled version of I_{TOT} called normalized I_{TOT}. Specifically, for each imaging parameter, we applied a normalization with respect to the maximum I_{TOT} value measured by varying the investigated parameter and maintaining the other three parameters constant [60]. It is important to highlight how the maximum I_{TOT} value was always measured when the empirical threshold was set to the minimum value, i.e., –37 dB, as a lower threshold includes a higher number of pixels in the computation of I_{TOT}.

Figure 7.3 shows the normalized I_{TOT} of the vertical artifact observed in the first scanning area (Area 1) as a function of center frequency. In this case, three different graphs are depicted, depending on the bandwidth (see the first three acquisition settings in Table 7.1). The three graphs (Fig. 7.3, top), representing acquisitions performed with bandwidths equal to 0.5 (blue), 1 (green), and 2 (red) MHz, are normalized with respect to the maximum value of I_{TOT} (measured at CF = 3 MHz and BW = 0.5 MHz, with an empirical threshold of I_{TOT} equal to -37 dB). This normalization of the graph allows us to also highlight the dependence of the normalized I_{TOT} on the bandwidth. The trend of the normalized I_{TOT} decreases as the frequency increases, regardless of the bandwidth. Moreover, considering the same center frequency, a wider bandwidth denotes a lower normalized I_{TOT}. The maximum value of normalized I_{TOT} is indeed observable when CF = 3 MHz and BW = 0.5 MHz (maximum = 0 dB, mean = –0.16 dB, minimum = –0.36 dB), whereas the minimum value is at CF = 6.5 MHz (maximum = –41.05 dB, mean = –52.80 dB, minimum = –64.93 dB). It is important to highlight how the error bars lengthen as the normalized I_{TOT} decreases. In fact, the difference between the empirical threshold values is increasingly impactful as the I_{TOT} decreases because the number of

pixels involved in the I_{TOT} computation is generally lower when the normalized I_{TOT} decreases.

Figure 7.4 shows the normalized I_{TOT} of the vertical artifact in Area 1 as a function

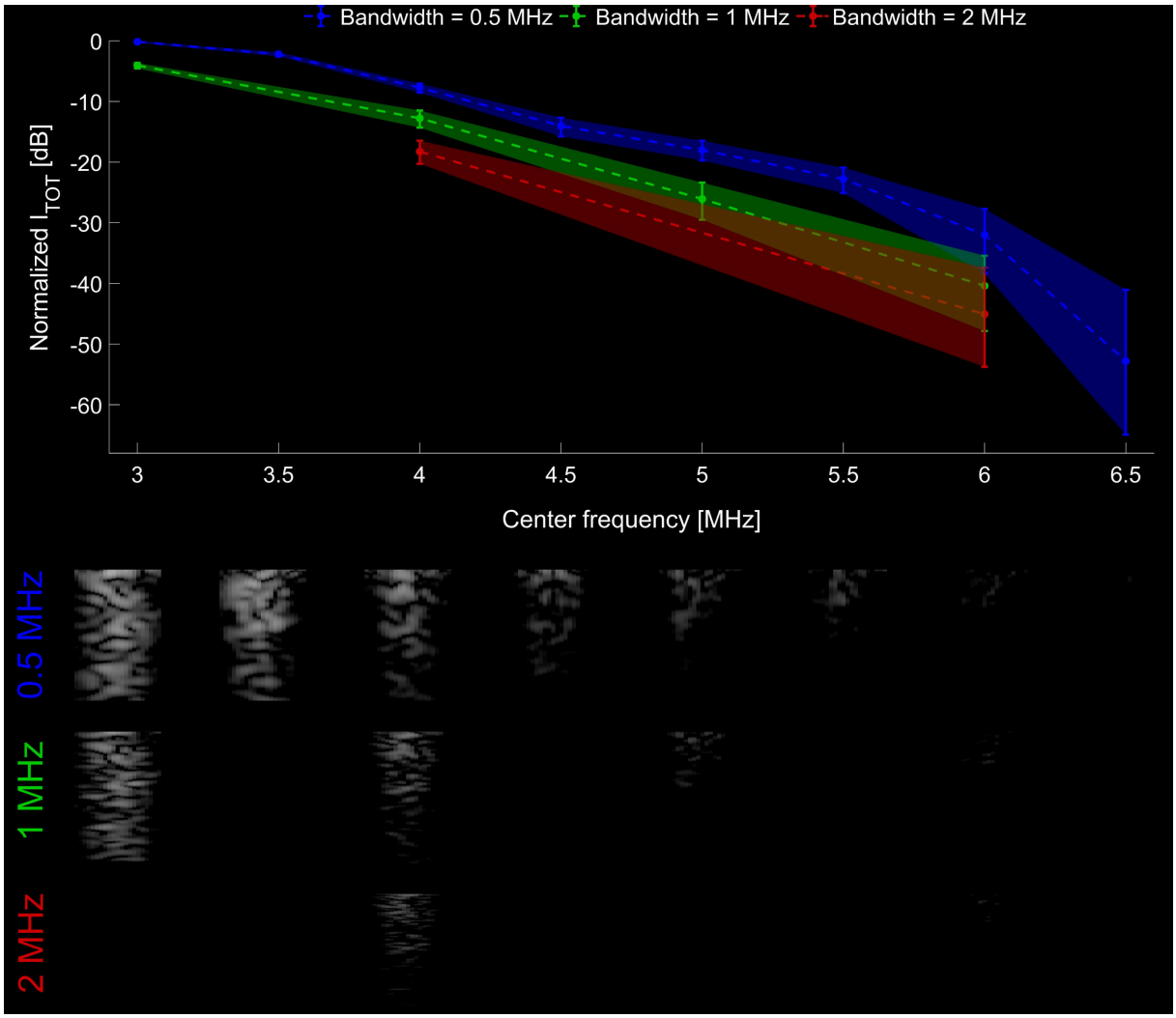

Figure 7.3: The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of center frequency (see top) is depicted together with the corresponding artifact (see bottom), shown with a 35-dB dynamic range. The points in the graphs represent the mean values of normalized I_{TOT} (computed by considering the five different empirical thresholds), whereas the error bars represent the minimum (computed with empirical threshold of I_{TOT} equal to -33 dB) and maximum (computed with empirical threshold of I_{TOT} equal to -37 dB) values of normalized I_{TOT} (computed by varying the empirical threshold from -37 to -33 dB with a 1-dB step size). The three graphs, representing acquisitions performed with bandwidths equal to 0.5 (blue), 1 (green), and 2 (red) MHz, are normalized with respect to the maximum value of I_{TOT} (measured at CF = 3 MHz and BW = 0.5 MHz, with empirical threshold of I_{TOT} equal to -37 dB). The focal point and the angle of incidence are fixed at 30 mm and 0°, respectively.

of the angle of incidence. The maximum value of normalized I_{TOT} is observed at AOI = –10° (maximum = 0 dB, mean = –1.85 dB, minimum = –3.94 dB), whereas the minimum value is at AOI = –30° (maximum = –4.14 dB, mean = –6.76 dB, minimum = –9.72 dB).

Figure 7.5 shows the normalized I_{TOT} of the vertical artifact in Area 1 as a function of bandwidth. As already seen in Fig. 7.3, the normalized I_{TOT} decreases as the bandwidth increases. Specifically, the maximum value of normalized I_{TOT} is observed when BW = 0.5 MHz (maximum = 0 dB, mean = –0.68 dB, minimum = –1.50 dB), whereas the minimum value is at BW = 2 MHz (maximum = –9.42 dB, mean = –11.22 dB, minimum = –13.24 dB).

Figure 7.6 shows the normalized I_{TOT} of the vertical artifact in Area 1 as a function of focal point. The normalized I_{TOT} increases as the focal point depth increases. This is particularly clear also by visually observing the vertical artifact (Fig. 7.6, bottom). In fact, when the focal point is 20 mm before the pleural line (i.e., FP = 10 mm) the vertical artifact is hardly visible, whereas its intensity increases when the focal point is set to 30 mm (i.e., the depth where the pleural line is located). By further increasing the focal point, the artifact remains clearly visible but widens laterally, most likely due to the effects of a deeper focus on the lateral resolution. The widening of the artifact is,

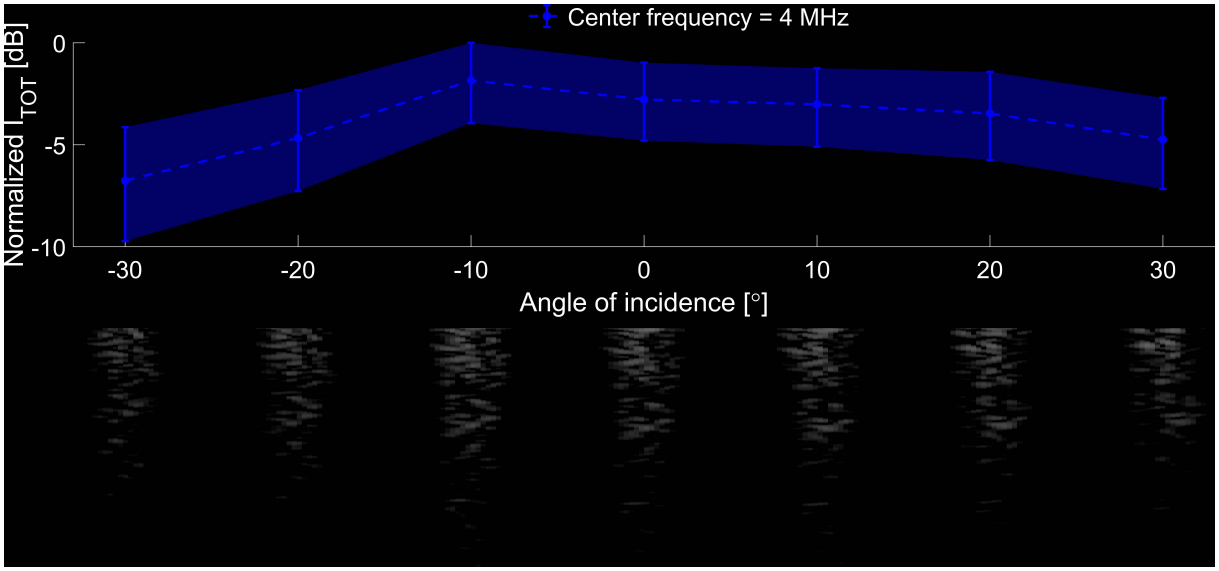

Figure 7.4: The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of angle of incidence (see top) is depicted together with the corresponding artifact (see bottom), shown with a 35-dB dynamic range. The points in the graphs represent the mean values of normalized I_{TOT} (computed by considering the five different empirical thresholds), whereas the error bars represent the minimum (computed with empirical threshold of I_{TOT} equal to -33 dB) and maximum (computed with empirical threshold of I_{TOT} equal to -37 dB) values of normalized I_{TOT} (computed by varying the empirical threshold from -37 to -33 dB with a 1-dB step size). The graph is normalized with respect to the maximum value of I_{TOT} (measured at AOI = –10°, with empirical threshold of I_{TOT} equal to -37 dB). The focal point, the bandwidth, and center frequency are fixed at 30 mm, 2 MHz, and 4 MHz, respectively.

in this case, translated in a higher normalized I_{TOT}. Specifically, the maximum value of normalized I_{TOT} is observed when FP = 50mm (maximum = 0 dB, mean = –2.09 dB, minimum = –4.65 dB), whereas the minimum value is at FP = 10mm (maximum = -16.59 dB, mean = -22.57 dB, minimum = –28.92 dB).

As shown in Fig. 7.7, the trend of normalized I_{TOT} as a function of the four different parameters is similar for all the nine investigated vertical artifacts. Specifically, the normalized I_{TOT} always decreases as the center frequency increases (Fig. 7.7, top) and, in some cases (see scanning areas 3, 5, 6, and 8), the vertical artifacts completely disappear (normalized I_{TOT} equals to -100 dB) at higher frequencies. The trend of normalized I_{TOT} as a function of angle of incidence (Fig. 7.7, bottom left) and bandwidth (Fig. 7.7, bottom center) does not significantly differ from the trend of the first vertical artifact (see Figs. 7.4 and 7.5; alternatively, see scanning area 1 in Fig. 7.7, bottom left and bottom center). As shown in Fig. 7.7 (bottom right), when evaluating the normalized I_{TOT} as a function of focal point, the highest values are observed when FP is set at the pleural line depth (30 mm) or deeper (40 and 50 mm).

Finally, to quantitatively evaluate the influence of the four parameters on the vertical artifacts' I_{TOT}, we have analyzed the I_{TOT} variation [60] by means of the boxplot shown in Fig. 7.8. Specifically, each box represents the I_{TOT} maximum variations obtained

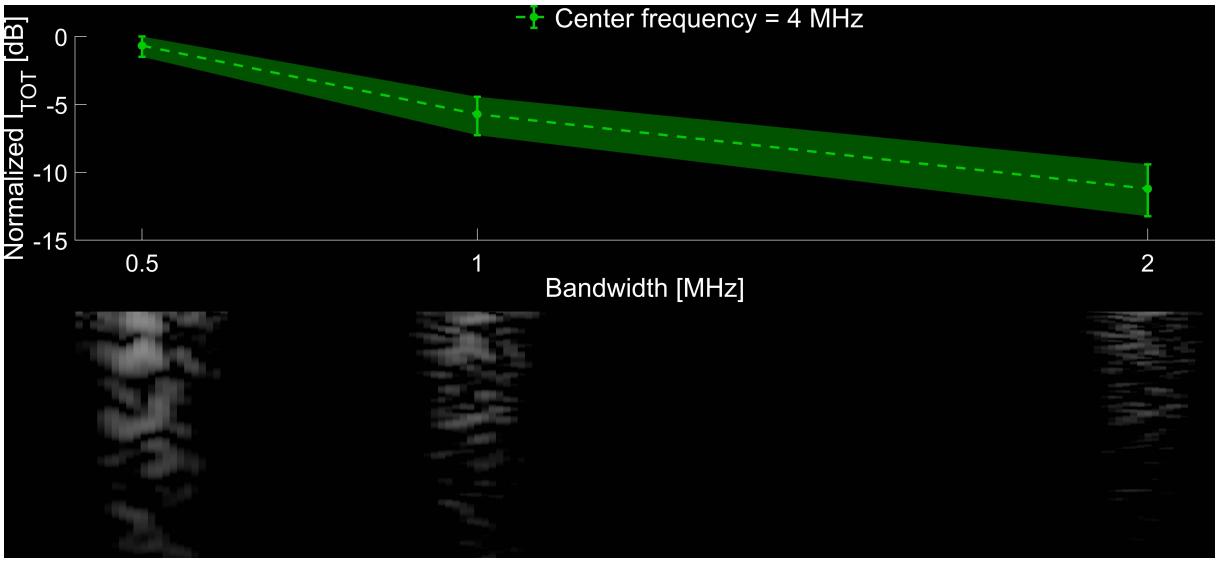

Figure 7.5: The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of bandwidth (see top) is depicted together with the corresponding artifact (see bottom), shown with a 35-dB dynamic range. The points in the graphs represent the mean values of normalized I_{TOT} (computed by considering the five different empirical thresholds), whereas the error bars represent the minimum (computed with empirical threshold of I_{TOT} equal to -33 dB) and maximum (computed with empirical threshold of I_{TOT} equal to -37 dB) values of normalized I_{TOT} (computed by varying the empirical threshold from -37 to -33 dB with a 1-dB step size). The graph is normalized with respect to the maximum value of I_{TOT} (measured at BW = 0.5 MHz, with empirical threshold of I_{TOT} equal to -37 dB). The focal point, the angle of incidence, and center frequency are fixed at 30 mm, 0°, and 4 MHz, respectively.

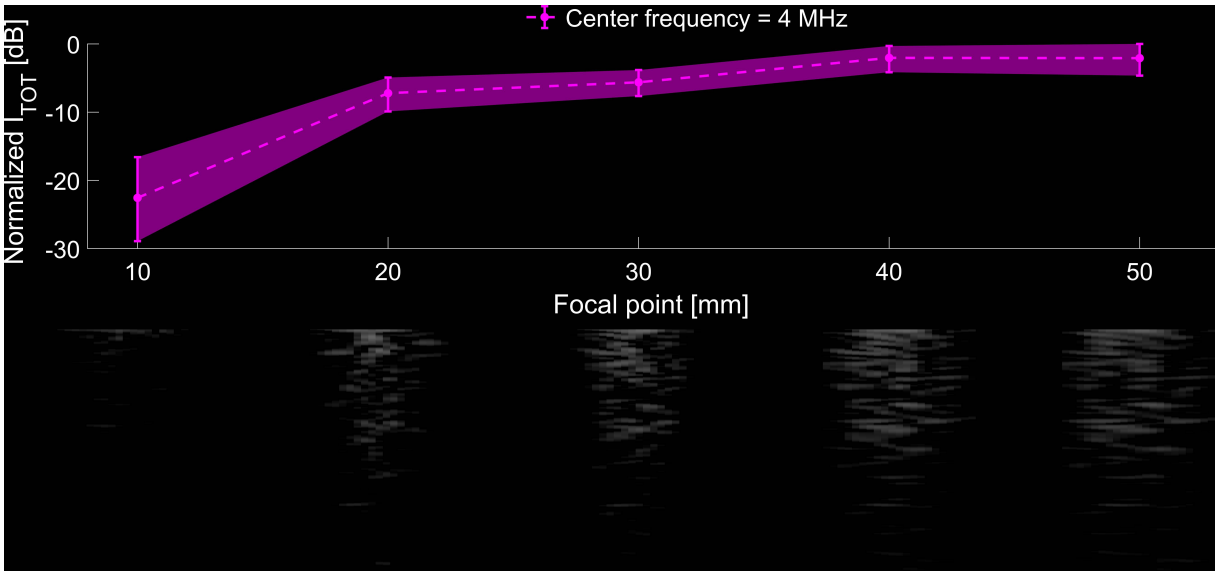

Figure 7.6: The normalized I_{TOT} of the vertical artifact observed in the Area 1 as a function of focal point (see top) is depicted together with the corresponding artifact (see bottom), shown with a 35-dB dynamic range. The points in the graphs represent the mean values of normalized I_{TOT} (computed by considering the five different empirical thresholds), whereas the error bars represent the minimum (computed with empirical threshold of I_{TOT} equal to -33 dB) and maximum (computed with empirical threshold of I_{TOT} equal to-37 dB) values of normalized I_{TOT} (computed by varying the empirical threshold from -37 to -33 dB with a 1-dB step size). The graph is normalized with respect to the maximum value of I_{TOT} (measured at FP = 50 mm, with empirical threshold of I_{TOT} equal to -37 dB). The angle of incidence, bandwidth, and center frequency are fixed at 0°, 2 MHz, and 4 MHz, respectively. The Water/Phantom's intercostal tissue interface (first interface) is at about 5-mm depth, whereas the Phantom's intercostal tissue/Phantom's pleura interface (second interface; pleural line) is at about 30-mm depth. The normalized I_{TOT} increases as the focal point depth increases. Higher values of normalized I_{TOT} can be observed starting from the pleural line depth (30 mm), as the structures to be imaged are located at pleural line.

by keeping fixed all the parameters except from the investigated one [60]. For example, when considering I_{TOT} variations caused by FP, we have applied the following steps to obtain the box shown in Fig. 7.8 (on the right). For each vertical artifact, the five I_{TOT} values obtained with empirical threshold = –35 dB (consistently with Fig. 7.7), CF = 4 MHz, BW = 2 MHz, AOI = 0°, and FP equal to 10, 20, 30, 40, 50 mm, respectively (see Table 7.1). Then, we computed the I_{TOT} maximum variation as the difference (within these five values) between the maximum and minimum I_{TOT}. Successively, we repeated this process for all the nine vertical artifacts, thus obtaining nine maximum variation values to analyze the impact of FP. These nine values were hence used to form the box related to FP (see Fig. 7.8, right). Figure 7.8 shows how the center frequency is the most impactful parameter (median value = 43.55 dB, interquartile range = 54.28 dB), followed by the focal point (median value = 22.38 dB, interquartile range = 16.21 dB), bandwidth (median value = 12.31 dB, interquartile range = 3.97 dB), and angle of incidence (median

value = 6.19 dB, interquartile range = 4.64 dB).

7.4 Discussion and Conclusions

LUS is currently widely used by clinicians to obtain important clinical information on the state of the lung surface. Nevertheless, LUS remains limited to qualitative or semi-quantitative analyses, as the implementation of quantitative approaches is nowadays slowed down by lack of dedicated hardware [49] and imaging modalities, as well as by

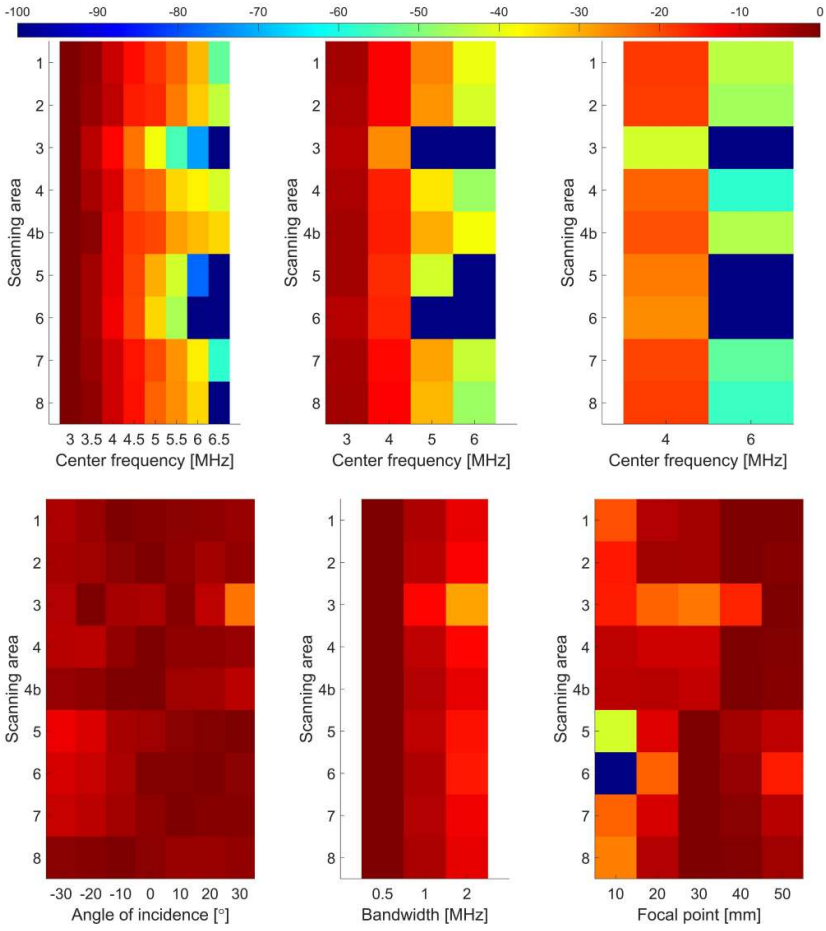

Figure 7.7: The normalized I_{TOT} of all the nine vertical artifacts, acquired from the eight scanning areas, are shown as a function of center frequency (top; BW = 0.5 MHz at top left, BW = 1 MHz at top center, BW = 2 MHz at top right), angle of incidence (bottom left), bandwidth (bottom center), and focal point (bottom right). The empirical threshold of I_{TOT} was set to -35 dB. For each group (CF, AOI, BW, FP), each row (scanning area) of cells was normalized with respect to its maximum value (as done from Figs. 7.3–7.6). By convention, a normalized I_{TOT} equal to -100 dB corresponds to the absence of the vertical artifact in the image. The normalized I_{TOT} values of scanning area 4 b refer to the second vertical artifact observed in scanning area 4.

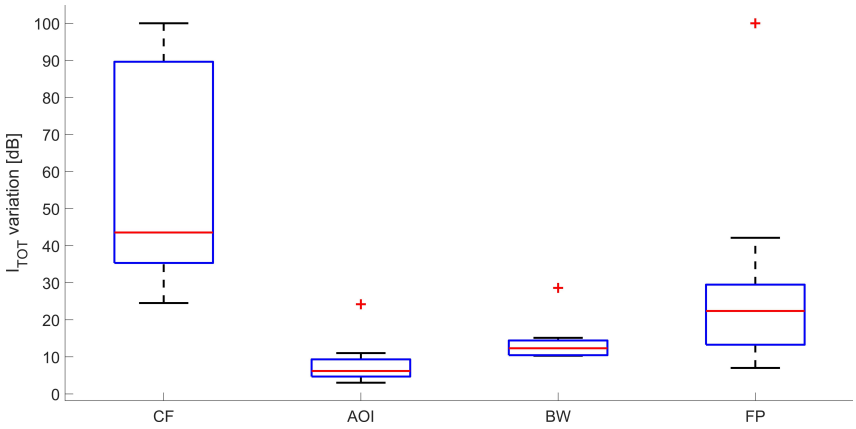

Figure 7.8: Boxplot showing the I_{TOT} variations caused by center frequency (first, left), angle of incidence (second), bandwidth (third), and focal point (fourth, right). Specifically, each box represents the I_{TOT} maximum variations obtained by keeping fixed all the parameters except from the investigated one. The red lines represent the median values, and the inferior and superior limits of the box are, respectively, the 25th and 75th percentiles, whereas the maxima and the minima are represented by horizontal black lines. The I_{TOT} variations were computed for each scanning area by setting an empirical threshold of I_{TOT} equal to -35 dB. This graph was hence derived from the values of Fig. 7.7.

the incomplete understanding of the vertical artifacts' genesis [24, 42]. However, recent studies have investigated the origin of LUS vertical artifacts [16, 55, 60, 85, 96, 101], leading to the formulation of a plausible hypothesis. The vertical artifacts are indeed likely caused by the formation, along the lung surface, of acoustic channels (or traps) filled with media acoustically similar to the intercostal tissue layers, a fact that allows for ultrasound to propagate within this traps [24, 50, 54, 87]. Then, the bouncing of the ultrasound wave within these channels, depending on the relation between the imaging frequency, the trap content, and geometrical property, can cause a resonance phenomenon [55, 60], which originates the vertical artifacts.

As the genesis of these vertical artifacts is likely caused by a resonance phenomenon [55, 60], recently proposed quantitative approaches are based on the frequency characterization of vertical artifacts [55, 58–60]. However, as shown in a recent *in vitro* study [60], even though the center frequency of the emitted ultrasound wave seems to be the most impactful imaging parameter in the vertical artifacts' characterization, there are other parameters that should be considered when implementing quantitative approaches or when standardizing the acquisition process in semi-quantitative approaches. This to verify the stability of the artifacts' intensity of key imaging parameters.

For this reason, in this article, we have quantitatively evaluated the dependence of vertical artifacts on four different important imaging parameters (center frequency, focal point, bandwidth, and angle of incidence).

As shown in Fig. 7.8, CF is confirmed as the most impactful parameter in the vertical artifacts' characterization (median value = 43.55 dB, interquartile range = 54.28 dB). Moreover, the normalized I_{TOT} decreases as CF increases and, thus, the highest values

of normalized I_{TOT} are always observed at the lowest frequencies (see Fig. 7.7, top). The high I_{TOT} of the artifacts at lower frequency could be caused by the presence of acoustic traps having large dimensions [58, 101], and thus be linked to the dimension of the lung damage. It is important to highlight how these results are linked to the specific phantom utilized, and thus to its internal structure. Therefore, when evaluating different pathological conditions, characterized by different alveolar distributions and geometries, the results in terms of total intensity as a function of center frequency and bandwidth could vary. As shown in the first clinical studies, these variations carry diagnostic information [58, 59].

Moreover, the reduction of the artifacts' total intensity with growing bandwidth is also coherent with the hypothesis that vertical artifacts are generated by resonance phenomena. Indeed, narrow-band ultrasound pulses are more suitable to activate such a type of phenomena compared to wideband pulses, as the energy is more concentrated around a specific frequency. In this sense, frequency and bandwidth should be seen as an inter-dependent factor, with a narrow bandwidth to be favored as it improves artifacts visualization as well as the accuracy of the artifacts' frequency characterization. These results are consistent with results obtained with different phantom types [55].

However, this hypothesis, coherent with reported *in vitro* [55, 60] and *in silico* [96] results, should be further verified with studies where the exact dimensions and content of the trap are known.

It is important to highlight how the overall results show significant variability for all the four parameters (the lowest I_{TOT} variation is observed for AOI, with a median value = 6.19 dB, and an interquartile range = 4.64 dB; see Fig. 7.8). As a consequence, the dependence of vertical artifacts on these parameters should be considered, especially when implementing a quantitative approach. This result is of fundamental importance also for the standardization of the acquisition process in semi-quantitative approaches. In fact, as the operator currently has very little control over these parameters, the comparison of data acquired with different scanners cannot be properly performed. As an example, the use of a convex over linear probe affects the angle of incidence of the ultrasound wave (due to its convex shape), thus potentially affecting the analysis. This would imply the necessity of applying a compensation process when utilizing a convex probe to reduce the effect of the angle of incidence on the vertical artifacts.

In this paper, we have mainly studied the dependence of vertical artifacts on total intensity as this parameter was already proven to be a relevant indicator in characterizing the lung surface of patients affected by different pulmonary pathologies [58, 59]. However, there are other features, as the speckle pattern, that could be studied to characterize the scatterers' geometrical properties.

In conclusion, even though semi-quantitative approaches supported by computer aided methods [2, 70] can partially solve the problem of subjectivity, which is intrinsic to LUS pattern evaluation, they cannot provide quantitative methods. The visualization of vertical artifacts is indeed dependent on the parameters investigated in this study and, thus, semi-quantitative approaches remain limited, as they do not provide control on the aforementioned parameters.

For these reasons, the development and implementation of quantitative LUS ap-

proaches need to be supported by custom hardware and dedicated approaches that allow the analysis of RF data and the possibility to tune the key parameters [49], which affect the vertical artifacts' intensity.

Chapter 8

Numerical study on lung ultrasound B-line formation as a function of imaging frequency and alveolar geometries

This Chapter[7] *presents an in silico study investigating the dependence of LUS vertical artifacts on alveolar size and geometry. Specifically, this study focused on the factors affecting the artifacts' formation by numerically simulating the ultrasound propagation within the lungs through the toolbox k-Wave. Since the main hypothesis behind the generation of vertical artifacts relies on multiple scattering phenomena occurring once acoustic channels open at the lung surface, the impact of changing alveolar size and spacing is of interest. The tested domain is of size 4 cm × 1.6 cm, the investigated frequencies vary from 1 to 5 MHz, and the explored alveolar diameters and spacing range from 100 to 400 μm and from 20 to 395 μm, respectively. Results show the strong and entangled relation among the wavelength, the domain geometries, and the artifact visualization, allowing for better understanding of propagation in such a complex medium and opening several possibilities for future studies.*

8.1 Introduction

Lung diseases are among the most widespread globally [130]. As an example, chronic obstructive pulmonary disease is the third primary cause of death worldwide [130]. Recently, the viral pneumonia caused by the severe acute respiratory syndrome coronavirus 2 infection has affected more than sixty million people, including a million and a half deaths [131]. To monitor and diagnose the lung pathologies, ultrasound represents a suitable alternative compared to chest x-ray and computed tomography, being ultrasound

[7] This Chapter appears in:
[J6] E. Peschiera, F. Mento, L. Demi, "Numerical study on lung ultrasound B-line formation as a function of imaging frequency and alveolar geometries," in *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2304–2311, April 2021.

more accessible, portable, cost-effective and radiation-free. Nevertheless, the acoustical properties of the lungs, composed by millions of aerated alveoli that complicate the ultrasound propagation within the lung, render lung ultrasound (LUS) a challenging task [88]. The main assumption behind standard ultrasound imaging, requiring similar transmission media from the acoustical viewpoint, is not fulfilled when applied to the lungs. The impedance mismatch at the lung surface is extremely large and, consequently, an acoustic wave experiences a reflection coefficient close to one [88]. For these reasons, LUS relies heavily on artifactual images rather than anatomical ones.

The clinical information provided by LUS is based on the different artifactual pattern generated by a healthy and a pathological organ [21, 24, 85, 88, 132, 133]. In case of a healthy lung, the lung volume is mostly occupied by air and the pleural-line behaves as a perfect reflector to ultrasound, impairing the propagation beyond it [50, 88]. In this scenario, horizontal artifacts (or A-lines) appear as a result of the multiple reflections of the transmitted pulse between the lung surface and the probe [88, 133]. The spacing between A-lines is equal to the distance between the pleural-line and the probe. On the other hand, a pathology affecting the lung typically causes a reduction of the volume occupied by air in favor of fluids (e.g., water and blood) or soft tissues [50, 88]. This decrease in the acoustic mismatch between the intercostal tissue and the lung surface creates acoustic channels accessible to ultrasound waves [50]. In this case, LUS images are characterized by the presence of vertical artifacts, or B-lines [16, 17]. The most accepted hypothesis behind their generation relies on the multiple scattering of the ultrasound pulse within acoustic traps below the lung surface, which behave as a secondary ultrasound source [50, 54, 101]. A correlation between B-lines visualization and several lung diseases has been observed for decades, and different studies have been proposed to study the B-lines formation processes depending on the alveolar features, as well as on the imaging parameters [21, 24, 55, 58–60, 85]. Despite this, a study investigating in a highly controlled manner the impact of the alveolar geometrical distribution was lacking.

Previous *in vitro* works have highlighted how the B-lines' visualization depends on the imaging frequency [55, 60]. Moreover, the frequency at which the spectral peak occurs, called native frequency, is suggested to be linked with the alveolar size and distribution [54, 55, 101]. Each acoustic trap is therefore characterized by its spatial features, which determine the frequency able to produce a vertical artifact through resonance phenomena [54, 101]. Hence, the native frequency could be exploited to quantitatively evaluate the state of the lungs, and could help to characterize different diseases [54, 55, 59, 101]. Indeed, the alterations of the sub-pleural volume vary with the pathology, and this is testified also by the different B-lines patterns that are observed in clinical images [21, 24, 50, 85]. The artifacts can be, for instance, modulated or non-modulated in intensity, and can have various lateral extensions [24, 50]. Furthermore, recent clinical studies have shown the possibility to *in vivo* characterize B-lines based on their frequency content, and how such characterization can positively impact on LUS specificity [58, 59]. With these premises, the aim of this research is to investigate, via numerical simulations, the B-lines formation by varying the imaging frequency and the features of the domain, i.e., the alveolar diameter and spacing. The frequency trend is of interest, as well as the impact of simulating lungs having different densities. In general, the implicit question of the work is whether

it is possible to obtain vertical artifacts with the numerical tool employed. In case of positive response, several future studies would open. Indeed, the numerical setup gives the complete freedom to define both the transmission part and the mimicked domains, as opposed to experiments made on phantoms where the level of control is obviously reduced.

In section 8.2, the choices underlying the numerical design are presented, from the imaging modality to the definition of the computational domains, as well as the methods used to evaluate the outcomes of the simulations. The obtained results are presented in section 8.3 and commented mainly in section 8.4, where future works are also proposed.

8.2 Methods

The numerical simulations have been performed through the k-Wave MATLAB toolbox [97]. The computational domains replicate the lung structure. Indeed, they are of size 4 cm $\times$ 1.6 cm, where the first two centimeters in depth are composed of muscle tissue, while air inclusions are introduced in the last two centimeters. Section 8.2.1 analyzes the main features of the transmission and reception phases, as well as of the spatial resolutions. The following section, 8.2.2, presents the numerical domains and the main parameters describing them. Lastly, section 8.2.3 introduces the methodology used to evaluate the intensity of the B-lines displayed in the reconstructed images.

8.2.1 Imaging parameters

The images are formed using a linear array beamforming modality, whose main parameters are listed in Table 8.1. Beam's focalization is implemented, with focal point placed at the pleural-line. The transducer is located at a depth of approximately 0.3 mm from the beginning of the computational domain. Being N the number of array elements and M the sub-aperture dimension, the number of lines in an image equals N-M+1=17. Every line is formed by fixing the number of sensors to M and moving each time the underlying domain, mimicking the displacement of the sub-aperture. The elements' excitation signals are modulated Gaussian pulses with central frequency and pulse length equal to f_0 and 2 μs, respectively.

At the end of the reception phase, where the same focalization as in the transmis-

Parameters	Values
Array elements (N)	80
Sub-aperture elements (M)	64
Pitch	245 μm
Kerf	45 μm
Focal depth	2 cm
Pulse bandwidth	1 MHz
Transmitted peak pressure	1 MPa

Table 8.1: Imaging array parameters, being the bandwidth computed at –6 dB from the spectral peak. The array aperture can be computed as dx·(Pitch·N-Kerf), where dx is the grid size.

f_0 (MHz)	1	2	3	4	5
dx (μm)	12.5	12.5	12.5	10.71	10
dt (ns)	0.4	0.4	0.4	0.34	0.32

Table 8.2: Domain grid sizes (dx) and time sampling intervals (dt) for each imaging frequency.

f_0 (MHz)	1	2	3	4	5
LR (mm)	3.10	1.61	1.10	0.84	0.66

Table 8.3: –12 dB lateral resolutions (LR) for each imaging frequency.

sion phase has been adopted, the envelope is extracted through Hilbert transform and conversion to logarithmic scale is performed. Once all the lines are formed, the image is normalized with respect to its maximum value and, representing power values, is visualized with a -40 dB of dynamic range. The studied frequencies vary from 1 to 5 MHz. The domain's grid size dx is computed considering the smallest wavelength occurring within the numerical domain, while the time sampling interval dt is defined by the simulator itself. Both quantities, that depend on f_0, are reported in Table 8.2.

The -12 dB axial resolution (AR) has been computed taking as reference the speed of sound within the muscle tissue (i.e., 1580 m/s), and equals AR=0.89 mm. The lateral resolutions, instead, are frequency-dependent and have been estimated from the beam profiles. Their values are listed in Table 8.3.

8.2.2 Computational domains

The computational domains are composed of two media, muscle medium (whose acoustic properties are c_0=1580 m/s and ρ_0=1041 kg/m^3) and air (characterized by c_0=300 m/s and ρ_0=1.23 kg/m^3). An example of domain is provided by Fig. 8.1, in which the image at the top depicts the entire computational area. The simulated lung surface is located at a depth of 2 cm, below which aerated medium is present. The lung area is subdivided into three parts along the lateral direction, two of which are entirely composed of air, while the central one includes both muscle tissue and alveoli. In the latter region, indicated by a red box in Fig. 8.1, deformations in the alveolar geometries are introduced to model different alterations of the lungs' structure. The internal lung structures are formed by a regular pattern formed by an alveolus and a spacing that repeats itself along both the dimensions. Consequently, the alveoli are aligned along the depth. This is clearly an oversimplified model of lung tissue. However, it allows for a detailed analysis on the impact of alveolar dimensions and spacing in a highly controllable fashion. The reconstructed images (see Fig. 8.2 and 8.8) will show how the proposed domains' configuration enables only the generation of a unique, relatively large B-line located at the center of the lateral axis. In a clinical situation, B-lines appear as isolated or coalescent vertical artifacts, the lateral extend of which depends on the extent and kind of alterations present at the lung surface [50, 88]. In order to study this variability, dedicated numerical domains should be investigated. In this study we opted instead to adopt a unique and regular pattern to

start unravelling the interconnections between alveolar diameter, alveolar spacing, imaging frequency and B-lines' intensity. An improvement of the domains' physical coherency with respect to the many conditions that may be found in a clinical situation would require introducing local irregularities in the features of the acoustic channels and assess their impact on the generated B-lines.

The parameters modelling the domains are the alveolar diameter D and spacing s, defined as the distance between the end of an air sphere and the beginning of the adjacent one (see Fig. 8.1). The examined diameters are 100, 200, 300, and 400 μm, while the investigated spacing correspond to 20, 100, 198, 263, and 395 μm. A healthy lung features a diameter of approximately 280 μm and a spacing of less than 10 μm [11]. Therefore, all the tested cases replicate pathological situations. The three largest values of spacing (i.e., 198, 263 and 395 μm) are equal to half of the wavelength (within the muscle medium) at 4, 3, and 2 MHz, respectively. Hence, the results associated to them are of interest to observe potential resonance phenomena, associated to constructive interference of scattered waves.

It is worth highlighting how the digitally approximated values of the alveolar parameters, indicated by D' and s', may differ from D and s because of numerical approximations

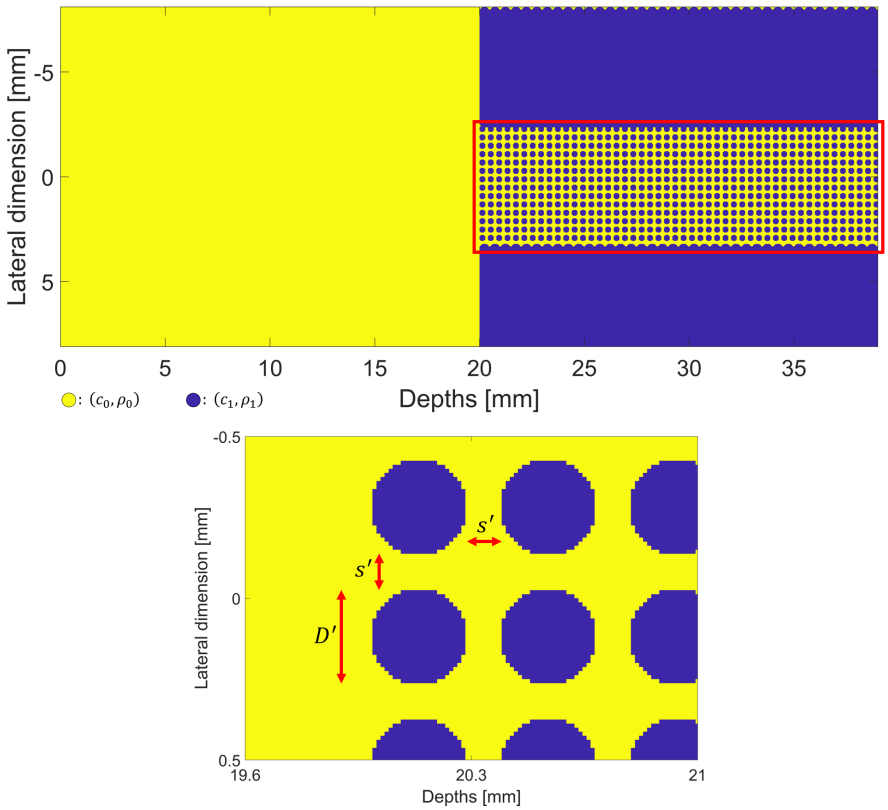

Figure 8.1: Example of numerical domain, complete view (top) and enlargement (bottom). Deformations in the alveolar size and distance are introduced in the region surrounded by the red box. The digitally approximated diameter D' and spacing s' are indicated. The couple (c_0, ρ_0) represents the acoustic properties of muscle, while (c_1, ρ_1) is associated with air.

D (μm)	s (μm)	D' (μm)	s' (μm)	$\overline{s'}$ (μm)
100	20	88	38	44
400	20	388	38	104
100	100	88	113	119
400	100	388	113	179

Table 8.4: Original parameters (D, s), digitally approximated counterparts (D', s'), and average channel spacing ($\overline{s'}$) of four numerical domains. As expected, both alveoli and spacing having larger sizes are better approximated, and $\overline{s'}$ differs considerably from s' in the case of large alveoli.

when converting the information from the physical (meters) to the numerical domain (pixels). Moreover, the spacing s' represents only the minimum distance between two alveoli but, due to their spherical shape, different spacing are present between two adjacent spheres. Calling s'_k the kth spacing and x_k its occurrence within a single acoustic channel, the average alveolar spacing can be computed as $\overline{s'} = (1/K_{tot}) \sum_{k=1}^{K} x_k s'_k$, where K_{tot} is the total number of spacing within a channel. The variables D', s' are useful to assess whether the original proportions on the spatial quantities are met within the numerical domains, while $\overline{s'}$ allows to characterize a transonic channel in its entirety. Table 8.4 provides some examples in this regard. Another useful parameter that can be extracted from the domains' analysis is the sub-pleural muscle percentage, denoted by p. It is computed, from the red box in Fig. 8.1, as the percentage of pixels composed of muscle within the central region below the lung surface and it reflects the density of the mimicked lung.

8.2.3 B-lines' intensity metrics

A single outcome of a simulation is represented by a reconstructed image, depicting normalized pressure power in dB scale, and visualized with a -40 dB of dynamic range. An example of image is given in Fig. 8.2, together with the corresponding central line signal.

A strong reflection at a depth of 20 mm, referred as the pleural peak, is observed in every image. Subsequently, a potential B-line artifact can be noticed, starting from a depth of approximately 21 mm and reaching the end of the domain. Being the domains deterministic, the region in which the vertical artifact can be formed is the same for every simulation. This region, called ROI_B and indicated in light blue in Fig. 8.2, extends from the 6th to the 12th line along the lateral direction, and from 20.8 mm to the end of the domain along the depth. The image lines that delimit ROI_B (i.e., 6th and 12th) have been selected empirically, evaluating where, in the reconstructed images, the B-lines were visualised when considering a -40 dB dynamic range. The starting depth (i.e., 20.8 mm) has been computed by adding to the depth of the pleural interface (i.e., 20 mm) half of the axial resolution (calculated considering the entire pulse length and equal to 1.6 mm). To quantify the B-line's strength, the total intensity parameter has been computed as [59, 60]

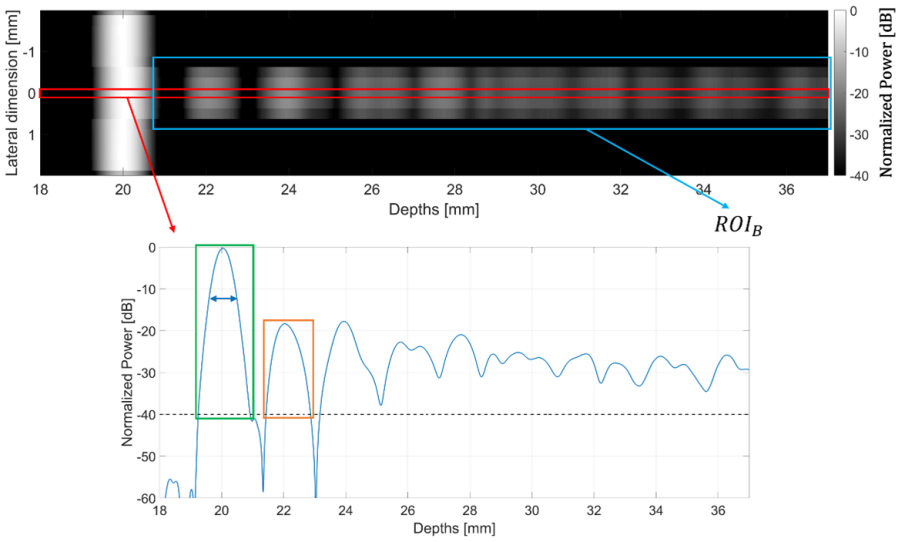

Figure 8.2: Example of reconstructed image (top) and corresponding central line (bottom), indicated in red. The image has been obtained employing f_0 = 3 MHz, D = 300 μm, and s = 100 μm. The B-line's region of interest, ROI_B, is highlighted in light blue. On the line signal, the green box indicates the pleural-line reflection, while the orange box indicates the first echo coming from within the lung. The second echo is arguably due to the scattering of the ultrasound wave between the first layers of alveoli. However, given the complexity of the multiple scattering phenomena in such a highly scattering environment, it is difficult to discuss this aspect with certainty. Simulations with a different number of layers, e.g., from 1 to 10, should be investigated to analyze this aspect, and these are the focus of future studies. The blue arrow shows the -12 dB pulse width. At 3 MHz, a pixel has physical sizes equal to the pitch along the lateral dimension (i.e., 245 μm) and to $dt \cdot c_0/2 = 0.31$ μm along the depths.

$$I_{tot} = 20 \log_{10}(A_{pix} \sum_{i,j} 10^{\frac{ROI_B(i,j)}{20}}) \tag{8.1}$$

where A_{pix} is the pixel area and $ROI_B(i,j)$ represents the image intensity in dB at the ith row and jth column in ROI_B, neglecting the values under -40 dB. The I_{tot} parameter evaluates a B-line considering both its intensity and its spatial extension (thanks to the summation) [59, 60]. Finally, the multiplication by A_{pix} considers possible differences in the grid sizes among the examined cases [59, 60].

After having computed the total intensity on all images (a total of R=100 results, considering all the combinations of parameters, was evaluated), normalization with respect to the maximum value in the whole set, I_{tot}^{max}, has been applied. Then, a threshold of -20 dB has been introduced to discard the images in which no B-lines are generated. Indicating with $I_{tot}(i)$ the non-normalized intensity of the ith image, the first step computes

$$\widehat{I_{tot}^{N}}(i) = I_{tot}(i) - I_{tot}^{max} \ [dB], \quad \forall i = 1, ..., R. \tag{8.2}$$

The second step, instead, applies a threshold as follows:

$$I_{tot}^{N}(i) = \begin{cases} \widehat{I_{tot}^{N}}(i)\ [dB], & \text{if } \widehat{I_{tot}^{N}}(i) > -20\ [dB] \\ -20\ [dB], & \text{if } \widehat{I_{tot}^{N}}(i) \leq -20\ [dB] \end{cases}, \quad \forall i = 1, ..., R. \tag{8.3}$$

8.3 Results

All the quantitative results (Figs. 8.3, 8.4, 8.5, 8.6, 8.7) show how, regardless of the sizes of alveolar diameter and spacing, at an imaging frequency of 1 MHz no visible B-line is produced. At 2 MHz, apart the images obtained with s=395 μm, only tenuous artifacts are produced. Fig. 8.3 depicts the I_{tot}^{N} values referring to a spacing of 20 μm. The frequency peaks are located at f_0=3 MHz and, in general, the alveoli of size D=300 μm are associated to the strongest artifacts. However, the most intense B-line is achieved when a frequency of 3 MHz is employed and scatterers having diameter of 100 μm are

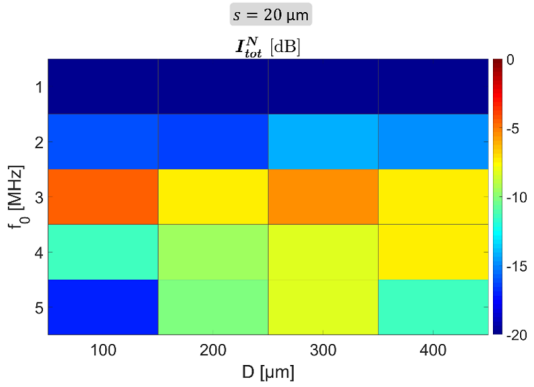

Figure 8.3: Normalized total intensity parameter as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing s = 20 μm.

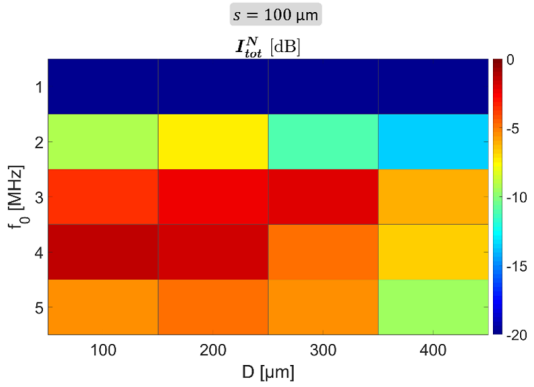

Figure 8.4: Normalized total intensity parameter as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing s = 100 μm.

replicated within the domain.

When the spacing is 100 μm (Fig. 8.4), a general increase in the artifacts' intensity is noticed. The frequency trend moves towards 4 MHz and, as it happens for the cases illustrated in Fig. 8.3, no monotonic relationship is observed between f_0 and the intensity parameter. Indeed, even though the increase of frequency from 1 to 3 MHz is always associated to a growth in the B-lines' strength, this turns out to be not true for frequencies greater or equal than 3 MHz. This phenomenon seems to confirm the nature of the vertical artifacts as generated from resonance phenomena (provided that f_0 is above a certain value). Moreover, Fig. 8.4 illustrates how the small alveoli (i.e., of diameter D=100 μm and D=200 μm) are linked to the more intense B-lines, as opposed to what happens when s=20 μm.

By analyzing Fig. 8.5, referring to the results at s=198 μm, the same considerations can be drawn. The frequencies achieving the more evident B-lines are 3 and 4 MHz, with a prevalence of 4 MHz. The trend with respect to the alveolar diameter is in favor of the smaller alveoli. Comparing Fig. 8.3, 4 and 5, it follows how denser lungs (composed of alveoli of small size) are associated to the strongest B-lines, provided that the spacing is greater than a certain quantity (in our case, 20 μm). It is of particular interest relating the intensity peaks along the frequency with the alveolar spacing. Indeed, the quantity s=198 μm equals half of the wavelength (within the muscle tissue) at 4 MHz, that is $\lambda = c_0/f_0$=1580 m $s^{-1}/(4 \times 10^6$ Hz) = 395 μm.

When the spacing assumes the value of 263 μm (Fig. 8.6), the frequency trend shifts towards lower frequencies. In particular, the most salient outcome is represented by the f_0 value at which the strongest B-lines are formed, i.e., 3 MHz. The wavelength at f_0=3 MHz, in fact, is equal to twice the considered spacing. The alveoli of small size continue, on average, to favor the generation of more intense artifacts.

The images obtained using the largest spacing (i.e., 395 μm) are associated to B-lines whose intensities are more similar along both the frequency and the diameter axes (see Fig. 8.7). Indeed, no sharp peaks are observed in the I^N_{tot} chart in Fig. 8.7 and the values are well spread. However, compared to the previous results, there is a tendency to

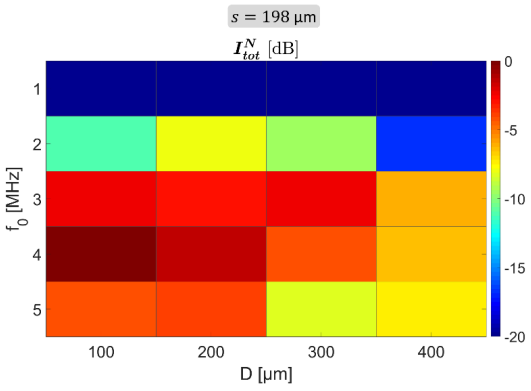

Figure 8.5: Normalized total intensity parameter as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing s = 198 μm.

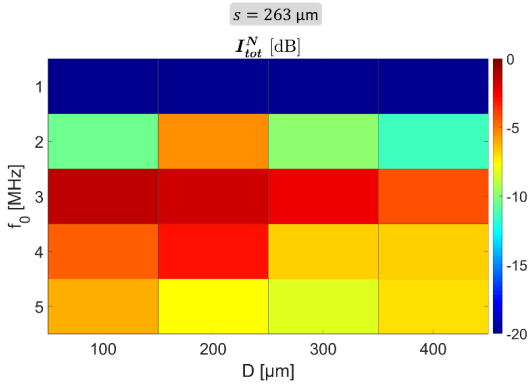

Figure 8.6: Normalized total intensity parameter as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 263$ μm.

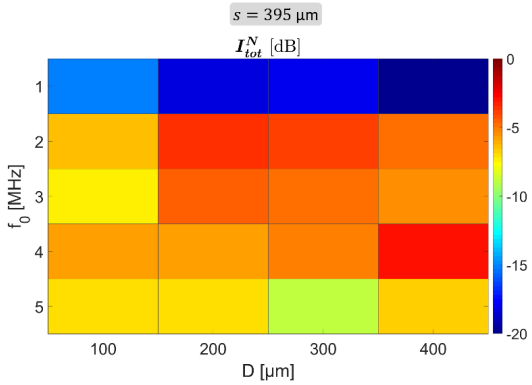

Figure 8.7: Normalized total intensity parameter as a function of alveolar diameter and imaging frequency, referring to the alveolar spacing $s = 395$ μm.

prefer the low frequencies, especially 2 MHz. Also in this case, the frequency that allows the formation of the strongest artifacts is the one whose wavelength is a multiple of the alveolar spacing (indeed, s equals to half of the wavelength at 2 MHz). It should be noted that in this case, given the investigated frequency range, we have two frequencies (2 and 4 MHz) whose wavelength is a multiple of the spacing. As regards the intensity parameter behavior in the diameter domain, the large alveoli are more frequently associated to intense B-lines.

To appreciate the visualization of the vertical artifact at the different imaging frequencies, Fig. 8.8 is proposed. The most intense B-line is achieved when f_0=3 MHz. In this case (represented by the third image of Fig. 8.8), the B-line is associated to an amplitude which tends to be more constant if compared with the other frequencies.

Another worthwhile analysis is represented by the I_{tot}^{N} dependency upon the sub-pleural muscle percentage p. Fig. 8.9 testifies how a denser lung, composed of alveoli of small size and characterized by high values of p, correlates with a strongest B-line. As proof, when p is above the sixty percent, essentially all the I_{tot}^{N} values are located within -12 dB

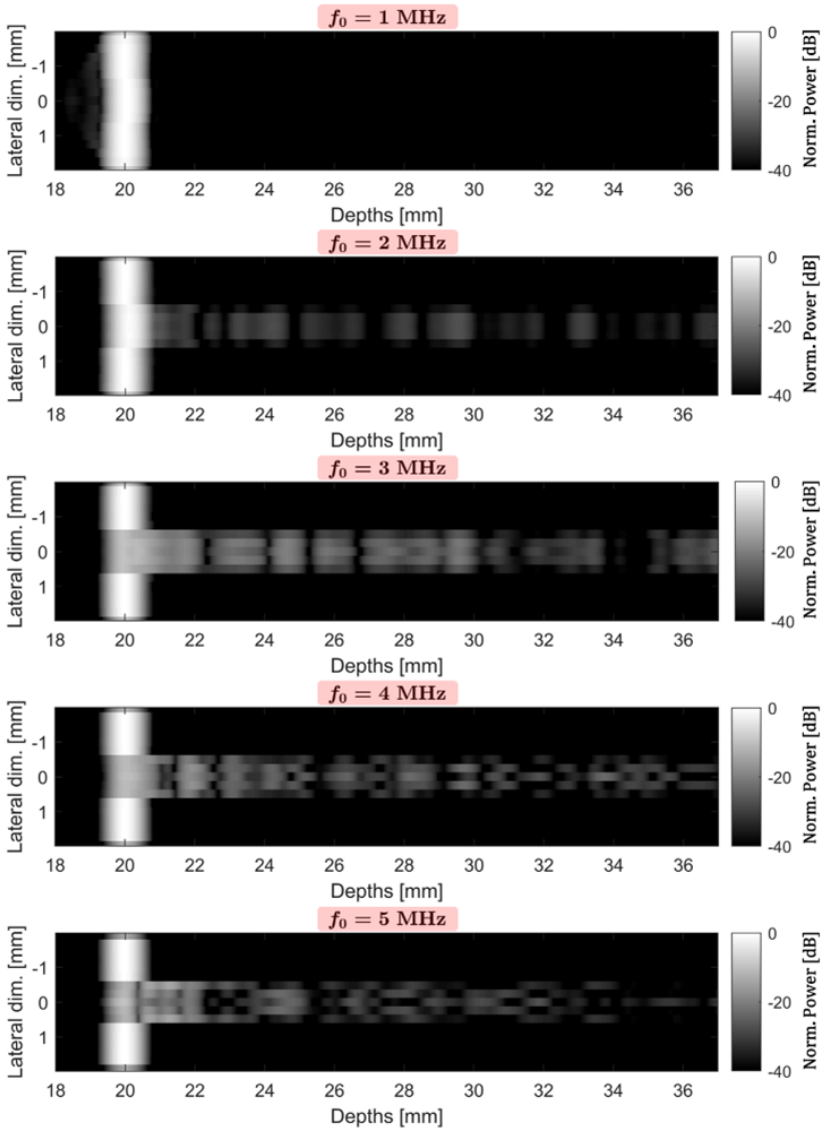

Figure 8.8: Reconstructed images at 1, 2, 3, 4, and 5 MHz (from top to bottom), referring to the numerical domain made of alveoli having diameter $D = 300$ μm and spacing $s = 263$ μm.

from the maximum intensity.

8.4 Discussion and Conclusions

Primarily, this work demonstrated the possibility to obtain B-lines artifacts by means of numerical simulations (through the k-Wave toolbox). Moreover, the results (Section 8.3) have shown a consistency with previous phantoms and clinical studies [55, 58, 60]. As an example, the behavior of the artifacts' formation in the frequency domain and with

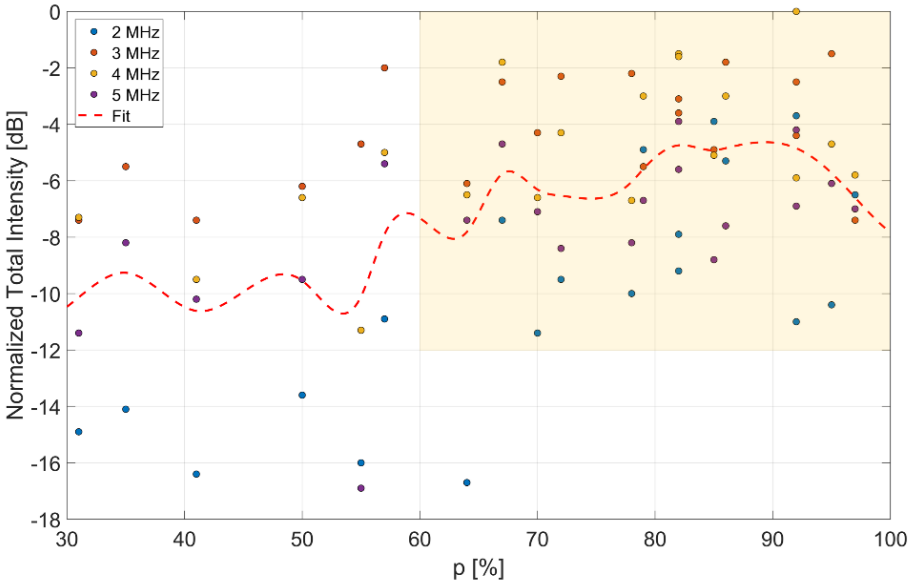

Figure 8.9: I_{tot}^{N} as a function of muscle percentage below the lung surface p. The 80 values obtained employing all the diameters and spacing and frequencies from 2 to 5 MHz are shown, together with an interpolated average (computed by means of a smoothing spline).

respect to the density of the lung exhibits common traits between the different study approaches. Therefore, *in silico* experiments can help in the characterization of the B-lines' dependency upon several factors and in performing tests that would be unfeasible *in vitro* or *in vivo*.

The quantitative analysis (Figs. 8.3, 8.4, 8.5, 8.6, 8.7) has shown how low imaging frequencies (i.e., $f_0 \leq$2 MHz) generally do not allow the generation of consistent artifacts, regardless of the alveolar diameter choice. Only when the alveolar spacing assumes the largest investigated value (i.e., s=395 μm) the lowest frequencies start to enable the B-lines' formation. In this direction, it is reasonable that large values of λ requires larger channels to create artifacts. These results further strengthen the hypothesis that the frequency characterization of vertical artifacts can be used as an indirect measure of the state of the lung, i.e., the lower the frequency at which B-lines are visualized, the larger the channels formed between the alveoli, the more severe the lung condition. Furthermore, when s=395 μm, the wavelength at 2 MHz (i.e., 790 μm) equals twice the spacing, and this may contribute to the increase in the artifacts' strength. From 3 MHz, instead, the B-lines' intensity trend with respect to f_0 ceases to be monotonic, and maxima in the I_{tot}^{N} quantity occur for both 3 and 4 MHz (dependently on the alveolar spacing employed). Therefore, the value of the imaging frequency does not a priori determine the artifact's intensity, but the latter will depend on how the wavelength will interact with the alveolar geometries (determined by the quantities D', s', and $\overline{s'}$). Following the same reasoning, also the lateral resolution (that is inversely proportional to f_0) is inadequate to predict by itself the intensity of the formed artifact.

As the results achieved at s=198, s=263, and s=395 μm suggest, there may be a

correlation between the B-lines' strength and the ratio between the wavelength (in our analysis, within the muscle tissue) and the spacing. Indeed, when s equals half of the wavelength at f_0, the image obtained transmitting at f_0 is more frequently associated to a more intense B-line. The reason behind these considerations may lie in the resonance phenomena (in other words, constructive interference) occurring within the acoustic traps below the lung surface, as hypothesized by previous works [54, 60, 101]. Therefore, specific imaging frequencies will activate certain traps according to their spatial features. In the first instance, the spatial features can be represented by s, but a more detailed characterization would certainly require more than one variable. For instance, a histogram describing all the different acoustic paths between single alveoli can be computed and used to differentiate between different lung's conditions [58].

As regards the I_{tot}^N trend with respect to the density of the simulated lung, large values of p (that describe denser lungs, formed by small air spheres) correlate with more intense B-lines, as evidenced by Fig. 8.9. Behind this finding there may be the greater sub-pleural propagation that a denser lung enables.

In general, this study strengthens the role of the total intensity parameter, alongside the corresponding frequency f_0, in quantifying the state of the lungs, reflecting both the quantities $\overline{s'}$ and p. These two can be the quantities describing each acoustic channel and trap, and are expected to vary with the pathology affecting the lung.

Further studies would include the additional validation of the numerical results through *in vitro* or *in vivo* experiments, replicating as closely as possible the conditions and the parameters presented in this study. On the other side, changes in the numerical model might be inserted to replicate more faithfully the real scenarios. For instance, the simulations could be performed in three dimensions instead of two, and other media (e.g., blood or water) could be included within the lung instead of muscle. Then, other alveolar distributions may be investigated (e.g., introducing some randomness in the alveolar positions and sizes), as well as alternative values of D, s and f_0. The ability to vary the alveolar diameter and spacing within the same numerical domain (e.g., by combining areas of large and close alveoli and areas of small and spaced alveoli) would be of interest. Also, the significance of the depth a B-line reaches may be investigated. The impact of the angle of insonification on the artifacts' generation processes could provide further considerations, as well as the minimum number of alveolar layers needed for the generation of a B-line. Additional tests could use an unfocused transmission (e.g., a plane wave) to estimate more broadly the impact of the lateral resolution on the artifacts' generation, and could study the effect of frequency dependent absorption.

All data resulting from the described simulations are available in Ref. [134].

Chapter 9

Quantitative Lung Ultrasound Spectroscopy Applied to the Diagnosis of Pulmonary Fibrosis: The First Clinical Study

This Chapter[8] *presents a quantitative clinical study aiming at differentiating patients with PF from patients with other lung diseases. Specifically, a previously developed multi-frequency ultrasound imaging technique was utilized to acquire ultrasound images from 26 selected patients. The multi-frequency imaging technique was implemented on the ULA-OP platform and an LA533 linear-array probe was utilized. RF data obtained at different imaging frequencies were acquired and processed in order to characterize vertical artifacts based on their frequency content. In particular, the artifacts' native frequencies (the frequency at which a vertical artifact exhibits the highest intensity) and bandwidth (the range of frequencies over which a vertical artifact shows intensities within -6 dB from its highest intensity), as well as artifact's intensity, were analyzed. The results showed how the analysis of these features allows the differentiation of fibrosis with a sensitivity and specificity equal to 92% and 92%, respectively. These promising results strongly motivate toward the extension of the clinical study, aiming at analyzing a larger cohort of patients and including a broader range of pathologies.*

9.1 Introduction

Respiratory diseases are among the main causes of death and disabilities in the world [135]. A well-know example is represented by the recent COVID-19 pandemic [136]. Another significant group of pulmonary diseases is represented by interstitial lung diseases (ILDs), which comprehend over 100 disorders of the lung that often lead to the fibrosis of

[8]This Chapter appears in:
[J7] F. Mento, G. Soldati, R. Prediletto, M. Demi, L. Demi, "Quantitative Lung Ultrasound Spectroscopy Applied to the Diagnosis of Pulmonary Fibrosis: The First Clinical Study," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2265-2273, November 2020.

this organ. Particularly, idiopathic pulmonary fibrosis (IPF), which represents about 30% of ILDs [137], is one of the most common forms of lung scarring and affects approximately 3 million people in the world [138]. However, since pulmonary fibrosis (PF) can derive also from known secondary causes (contrary to IPF, which is a subgroup of PF whose causes are unknown), the overall number of patients affected by PF is likely higher. This disease is generally characterized by a poor prognosis, also due to the delayed diagnosis, which is particularly common in IPF [139]. Therefore, to begin early treatment and slow down the progression of the disease, a timely diagnosis is fundamental. The high-resolution computed tomography (HRCT) represents the most utilized imaging instrument to derive a PF diagnosis [138, 139]. Nevertheless, the availability of a faster and more accessible preliminary instrumental examination would be extremely useful to better and more rapidly evaluate the state of the lung and direct the patients to the most suitable clinical path.

Lung ultrasound (LUS) currently represents an exploitable resource that is potentially able to satisfy the aforementioned requirements [24]. LUS is indeed a cost-effective, safe, and portable technology that allows real-time imaging. Its clinical relevance has been receiving growing attention only since 1997 [17]. However, the presence of air, whose acoustic characteristics significantly differs from the soft tissues [140], strongly complicates the use of ultrasound technologies. Standard ultrasound imaging is indeed designed to anatomically investigate the human body by assuming a quasi-homogeneous speed of sound. Furthermore, the acoustic impedances of soft tissues and air are extremely different, hence creating an interface (pleural line) having a high reflection coefficient [50, 140]. For this reason, the healthy lung behaves essentially as a perfect reflector. As a consequence, ultrasound waves bounce multiple times between two strong reflectors, i.e., the pleural line and the probe, hence generating the horizontal artifacts known as A-lines [24]. In contrast, when the lung becomes pathological, vertical artifacts called B-lines appear in the reconstructed image. The correlation between the appearance of B-lines and numerous pathological conditions of the lung are largely mentioned in the literature [17, 21, 50, 108, 122, 141]. These pathological conditions are characterized by the partial replacement of the lung volume previously occupied by air with fluids or tissue, consequently decreasing the acoustic impedance mismatch between soft tissues and the lung surface [50, 140]. Particularly, this replacement causes the formation of acoustic channels (or traps) made of media, such as tissue, water, or blood, where ultrasound can propagate, thus generating the aforementioned vertical artifacts [17, 24, 50, 87].

Hypothesis on the origin of B-lines, and their link to the presence of acoustic traps, has been discussed since similar artifacts were observed in a bubbly medium generated by shaking soapy water [16], rigid structures surrounded by a homogeneous medium [101], foamy phantoms [85], and bubbly media [55]. These studies demonstrate that these artifacts cannot be only associated with the presence of very specific anatomical structures of the lung (e.g., interlobular septa). On the other hand, both the pathological lung and the phantoms described in the previously cited literature have in common the presence of acoustic traps. Understanding and characterizing the interaction between ultrasound waves and these traps could explain the generation of B-lines. Indeed, after having partially trapped the energy of the emitted acoustic waves, these traps probably act as secondary ultrasound sources, which gradually reradiate the energy to the probe [101].

Nonetheless, the genesis of B-lines still remains unclear [23]. Consequently, diagnoses today are only based on the visual interpretation of these artifacts and are, thus, subjective and qualitative [50, 55]. The so-called semiquantitative approaches are mainly characterized by counting the vertical artifacts which appear in the image and are particularly spread in the clinical world [123–125]. Moreover, the visual recognition of specific ultrasound patterns and the consequent introduction of scoring systems for evaluating the state of the lung have been recently introduced also in the context of the COVID-19 pandemic [1, 32]. Notwithstanding the clear potentiality of LUS in investigating the lung, these approaches remain strictly dependent on the subjective evaluation of physicians.

Therefore, to provide clinicians with objective information, a quantitative characterization of B-lines is needed. Recent studies indeed focused on developing ultrasound imaging techniques designed for the lung, which aim at providing a quantitative measure describing the state of the organ [55–57]. While the study by Zhang *et al.* [57] was based on measuring the surface wave speed on the lung, Demi *et al.* [55] and Mohanty *et al.* [56] investigated the relationship between the multiple reflections of ultrasound waves occurring within the acoustic traps and the signal received by the transducer [88]. Of particular interest is the characterization of B-lines with respect to their frequency spectrum [55], which could provide an indirect estimation of both size and shape of acoustic traps [54].

PF represents a disease where the alveolar disposition is heterogeneous and completely subverted with respect to its normal structure [54]. Therefore, the acoustic traps are irregularly distributed and have complex and various shapes, differently from other diseases, such as acute cardiogenic pulmonary edema, where the original alveolar distribution is almost entirely preserved [54]. These irregularities in the pleural plane could be potentially characterized by a frequency analysis of B-lines, whose spectrum is expected to have a broad bandwidth when generated by heterogeneously sized and distributed traps such as with PF [101].

In this article, the frequency content of the B-lines observed in 26 patients with lung pathologies is quantitatively evaluated by exploiting the multi-frequency approach presented in [55]. In particular, this research investigates the potentiality to discriminate patients affected by PF by exploiting the frequency dependence of B-lines, which is quantified with a new parameter called *"total intensity"*. This parameter provides a measure of the B-line strength independently of the transmitted pressure amplitudes (assuming that no saturation occurs), which are generally different depending on the transmission parameters (e.g., number of transmitting elements, focal point, and center frequency of the emitted pulse).

This article is organized as follows. The utilized equipment and acquisition methods are presented in Section 9.2.1, whereas Section 9.2.2 describes the procedure used to extract the total intensity parameter. Section 9.2.3 introduces the utilized features and their statistical analysis, whereas Section 9.2.4 presents the exploited classifiers. Then, the main results are shown in Section 9.3, and the conclusions are presented with the discussion in Section 9.4.

9.2 Methods

9.2.1 Data acquisition

A multi-frequency approach based on orthogonal subbands centered at different center frequencies (3, 4, 5, and 6 MHz) was implemented with an ULtrasound Advanced Open Platform (ULA-OP) programmable platform [126], which was used together with an LA533 (Esaote, Florence, Italy) linear-array probe to acquire raw RF data from 26 patients. This specific set of frequencies was adopted following the results obtained from a study on controlled lung-mimicking phantoms, which showed how analyzing B-lines within this frequency range allowed discriminating between phantom types made to model different levels of lung alteration [55, 88]. The utilized probe has a -6- and -12-dB bandwidth from 3.8 to 12 MHz and from 3.2 to 13.2 MHz, respectively. The maximum of the transducer transfer function is at 8 MHz. A Gaussian pulse having a 2-μs time length was transmitted for every center frequency, guaranteeing each subband to have a narrow bandwidth, i.e., 1 MHz at -10 dB. The amplitude of each pulse was properly defined with respect to the probe bandwidth in order to equalize the pressure outputs. Adopting a strategy based on the transmission of pulses with a narrow bandwidth is indeed fundamental to better detect B-lines and precisely evaluate their frequency spectrum [55]. A dynamic sinc apodization function was utilized. Focal point and maximum imaging depth were set at 2 and 6 cm, respectively. Given that the pleural line generally sits at a depth of about 2 ± 1 cm, this choice allowed the analysis of artifacts for depths up to at least twice the depth of the pleural line. The ultrasound signals were transmitted and received by means of a subaperture of 64 elements, which was linearly shifted over the entire array to reconstruct an image composed of 129 lines. This process was repeated sequentially alternating between the different center frequencies. A pulse repetition frequency (PRF) equal to 4 kHz was utilized. Each image was, thus, formed at a frame rate of PRF/129 = 31 Hz, although a frame rate of 7.75 Hz was obtained when considering the images that belong to the same frequency.

Given the imaging settings described above, the -12-dB lateral resolution measured at focus by means of a wire phantom equals 0.81, 0.66, 0.65, and 0.54 mm when imaging at 3, 4, 5, and 6 MHz, respectively.

The received signals were digitized at a 50-MHz sampling frequency. To increase the signal-to-noise ratio (SNR), the time gain compensation (TGC) was experimentally set by considering the dependence of attenuation on both depth and frequency [129]. A value equal to 1.5 dB/MHz cm was adopted.

The examined patients represent a subgroup of a larger population of 101 patients, who signed an informed consent form to participate in this study (approved by the ethics committee for clinical studies: study number 1089—comitato etico sperimentazione clinica CEAVNO). Subjects consecutively admitted to the Pulmonary Medicine Unit of the Gabriele Monasterio Tuscany Foundation of Pisa in the period November 2017–December 2018 were evaluated. No selection was made for pathology. Subjects younger than 18 years were excluded. During their stay, all subjects underwent thoracic ultrasound by a single operator (GS) with more than ten years of chest ultrasound experience. The patients were not requested to hold their breath during the ultrasound investigation.

Moreover, all patients were investigated for their symptoms according to the procedures usually used in the institution when indicated, without any restriction, at the discretion of the clinician (RP). Instrumental examinations, in addition to the physical examination, were standard radiographs, computed tomography, radioisotopic investigations, and cardiac catheterization. Thoracic ultrasound was performed blindly, and the operator who performed the ultrasound examinations was unaware of the clinical or anamnestic data of the patients. Patients had posterior, basal, paravertebral, and apical scans in a sitting position. In the supine position, front and side scans on all explorable areas, including supraclavicular areas, were acquired. Video clips and RF data of the significant findings were stored.

In this work, to compose a balanced data set, we selected all the patients diagnosed with IPF and an equal number of randomly selected patients without lung fibrosis. Therefore, the investigated subgroup consists of 13 patients (ten males and three females, with age ranging from 56 to 89 years and the mean equal to 76.8) suffering from IPF (Patient ID from 14 to 26) and 13 patients with other diseases (eight males and five females, with age ranging from 55 to 85 years and the mean equal to 74.9), such as chronic obstructive pulmonary disease (COPD) (Patient ID 1, 4, 5, 6, 7, 8, 9, 10, and 13), emphysema (Patient ID 2), pneumonia (Patient ID 3, 4, 7, 13), pulmonary hypertension (Patient ID 8 and 11), and allergic asthma (Patient ID 12). Some of the nonfibrotic patients suffered from multiple diseases. Standard clinical examinations (e.g., CT) were performed to derive these diagnoses.

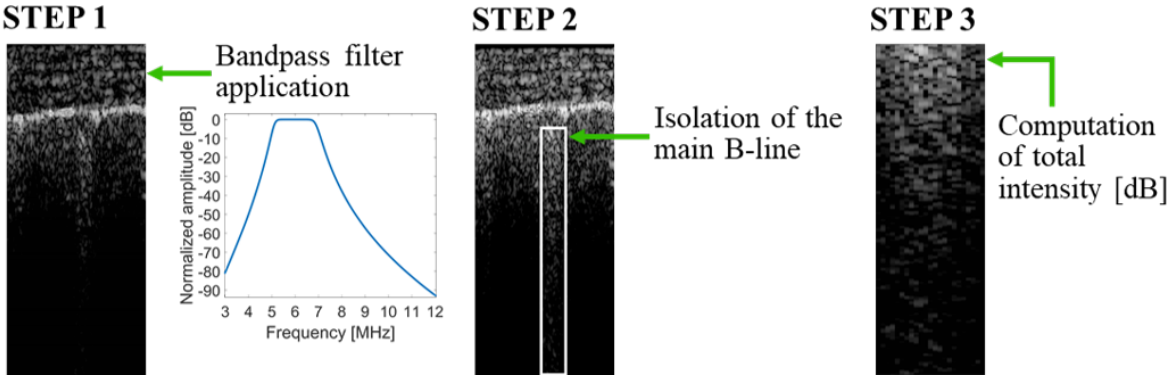

Figure 9.1: Procedure applied to the ultrasound data to compute the total intensity of B-lines. In step 1, a sixth-order bandpass Butterworth filter centered at the transmitted center frequency and having a 1.8-MHz bandwidth was applied. In step 2, the obtained image was displayed in logarithmic scale with a 35-dB dynamic range to manually segment the clearest B-line by defining a specific region of interest (ROI_B), represented by a white box in the figure (center). In step 3, the total intensity parameter was computed in ROI_B to quantitatively evaluate the B-line strength.

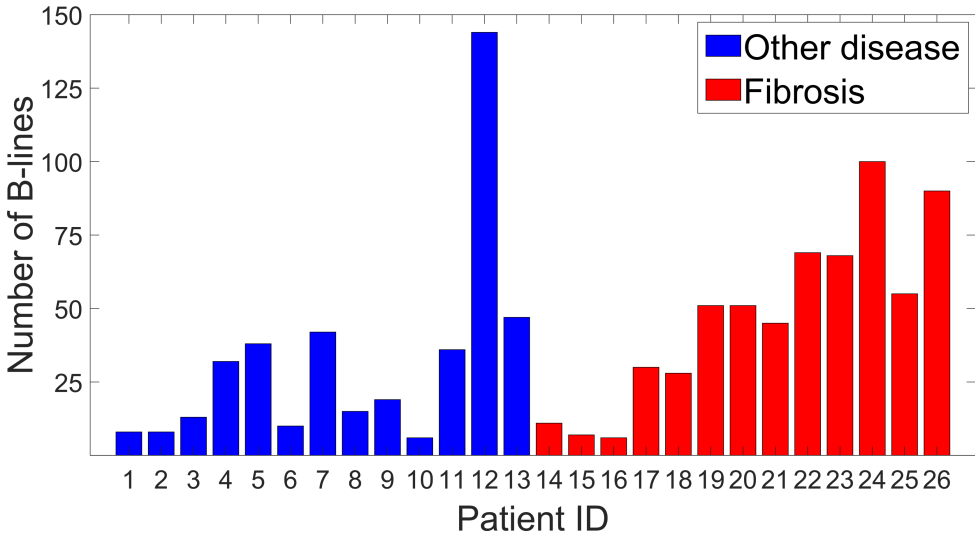

Figure 9.2: Number of B-lines for each patient. The patients with ID from 1 to 13 (blue bars) are diagnosed as nonfibrotic, whereas the patients with ID from 14 to 26 (red bars) are diagnosed as fibrotic.

9.2.2 Quantification of B-Lines' Intensity

B-lines' intensity was quantitatively evaluated by exploiting the three-step procedure depicted in Fig. 9.1. In step 1, to minimize the contributes due to unwanted frequency components, especially the harmonics, a sixth-order bandpass Butterworth filter centered at the transmitted center frequency and having a 1.8-MHz bandwidth was applied. This value was empirically chosen as a trade-off aimed at minimizing pulse distortion and maximizing noise removal. Moreover, the Hilbert transform was applied to extract the envelope, and the image was normalized with respect to its maximum value (displayed at the pleural line). In this way, the variability introduced by different thickness of intercostal layers on the relative amplitude associated with each B-line artifact (which starts at the pleural line) is mitigated. In step 2, the obtained image was displayed in logarithmic scale with a 35-dB dynamic range to visually detect the clearest B-lines over the different imaging frequencies. Each B-line was then manually segmented by defining a specific region of interest (ROI_B). Given a consecutive block of four images obtained at the four different frequencies, the ROI was manually selected from the image with the largely extended B-line, starting from 3 mm below the pleural line and in order to contain the entire artifact as visualized from a normalized image with a dynamic range of 35 dB. The same ROI was then utilized for the other three images of the block. In the final step, the total intensity parameter was used to quantitatively evaluate the strength of the selected B-lines. First, an arbitrary threshold was set at -35 dB, hence considering only the values above the threshold in the computation of the total intensity. This choice was made based on the noise level observed with the data. Then, the total intensity in logarithmic scale was obtained as follows:

$$I_{TOT} = 20\log_{10}(A_{pix}\sum_{i,j} 10^{\frac{ROI_B(i,j)}{20}}) \tag{9.1}$$

with A_{pix} being the pixel area (equal to 3.6×10^{-3} mm^2), i and j are the indexes referring to the pixel located in ROI_B at the ith row and jth column, and $ROI_B(i,j)$ represents the intensity (in decibels) of the pixel at the ith row and jth column.

Hence, this new metric introduced the possibility to quantify the relevance of a given B-line by considering both the intensity of the artifact and its spatial extension (implicitly included in the total intensity computation by summing all the pixels contained in ROI_B). Furthermore, thanks to the normalization operation (step 1), the dependence of the total intensity from the received signal amplitudes is mitigated, hence increasing the robustness and the reliability of the analysis.

A total number of 5,980 frames (2,944 frames acquired from the nonfibrotic group and 3,036 frames from the fibrotic group) was processed by means of the aforementioned procedure, leading to the detection of 1,029 B-lines.

9.2.3 Features and Statistical Analysis

In this study, we have considered three features, namely the B-line native frequency [55], bandwidth, and total intensity. The first two features correspond to the frequency at which a B-line exhibits the highest intensity and the range of frequencies over which a B-line shows intensities within -6 dB from its maximum (when the B-line's intensity is higher than -6 dB for all the investigated frequencies, the bandwidth equals to 4 MHz). The total intensity is introduced by 9.1. For each patient, multiple B-lines were analyzed with respect to these three features.

To represent each patient as a function of native frequency and bandwidth of its B-lines, Gaussian distributions were generated. Specifically, we computed the mean value and the covariance matrix of the native frequency and bandwidth of the B-lines observed in each patient. Successively, for visualization purposes, the covariance matrix was scaled by a factor of 10. A distribution of multivariate normal random numbers was then generated for each patient by means of the MATLAB function *mvnrnd*, given the mean values of native frequency and bandwidth, and the covariance matrix as inputs. Finally, a normalized multivariate normal probability density function was generated in MATLAB by means of the *mvnpdf* function, given the previously computed distribution of multivariate normal random numbers, the mean values of native frequency and bandwidth, and the covariance matrix as inputs.

To evaluate the performance of classifiers based on the three features introduced in this article, each patient was characterized with the mean values of native frequency, bandwidth, and total intensity of its B-lines (the average values of the B-lines observed in the patient). We referred to these parameters as *"mean native frequency,"* *"mean bandwidth"*, and *"mean total intensity"*.

To evaluate the statistical significance of each feature, an analysis of variance (ANOVA) test was performed between the fibrotic and nonfibrotic groups. In particular, the mean parameters were investigated.

The ANOVA tests were performed by means of the MATLAB function *anova1*.

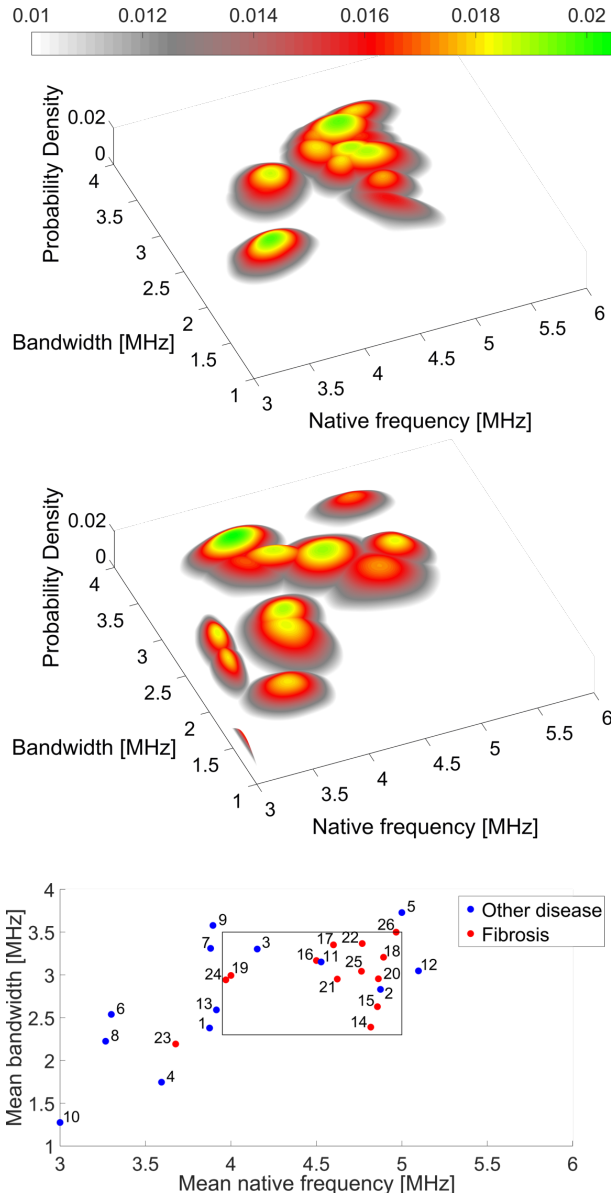

Figure 9.3: Gaussian distributions representing the 13 fibrotic patients (top) and the 13 non-fibrotic patients (center). The Gaussian distributions were generated by considering the mean native frequency, the mean bandwidth, and the covariance matrix of native frequency and bandwidth of B-line artifacts. The covariance matrix was divided by a factor of 10 to improve the visualization. The 26 patients are jointly depicted in the 2-D plot (bottom), where the x-axis and the y-axis represent the mean native frequency and the mean bandwidth of the B-lines observed in each patient, respectively. The patient IDs are written near the corresponding points. The patients within the black box are classified as fibrotic by the empirically defined binary classifier.

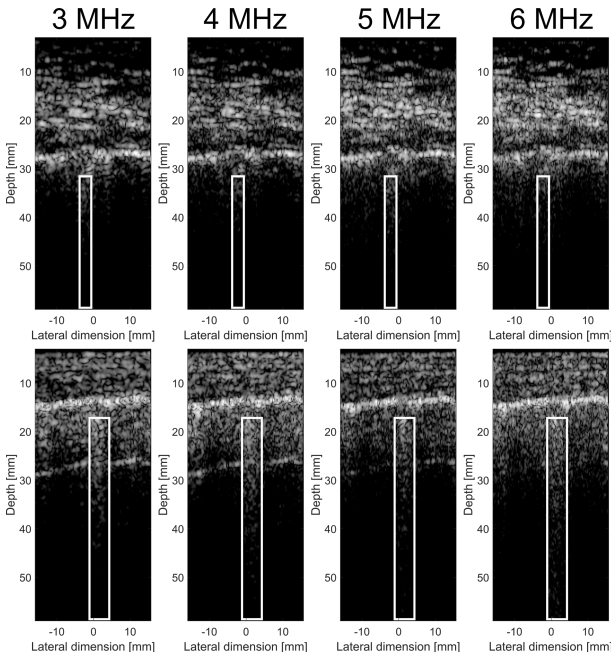

Figure 9.4: Multi-frequency images of the fibrotic patient labeled with ID 16 (top) and the nonfibrotic patient labeled with ID 7 (bottom). The regions of interest (ROI_B) that were used to compute the total intensity are represented as white boxes. The images are shown with a 35-dB dynamic range.

9.2.4 Classifiers

To objectively evaluate the potentiality of the mean native frequency, mean bandwidth, and mean total intensity for discriminating fibrotic patients, we trained and tested two different well-known classifiers, i.e., support vector machines (SVMs) with Gaussian kernels [142] and decision trees [143]. More specifically, both the SVMs and decision trees classifiers were implemented in MATLAB by means of the *fitcsvm* and *fitctree* functions, respectively. Each classifier was trained and tested by using k-fold cross validation, with k ranging from 2 to 26, where 26 is the number of samples (patients) in the data set. Since k-fold cross validation strictly depends on the randomly chosen groups, we performed a repeated k-fold cross validation by repeating 20 times the cross validation process for each value of k. Then, the performance results (accuracy, and specificity and sensitivity with respect to the fibrotic group) were averaged to obtain more objective measures of performance. To evaluate the classifiers' performance by testing all the possible combinations of the aforementioned features, the process was repeated seven times, i.e., by testing each feature separately, each couple of features, and the three features together. Finally, an empirical evaluation with respect to the classifiers' parameters was performed, and the SVM (*BoxConstraint* and *KernelScale*, set to 10 and 1, respectively) and the decision tree (*MinParentSize* set to 10) having the best accuracy were chosen. Moreover, an empirically defined binary classifier was introduced to show the best performance achievable with this data set. This classifier was defined by visually observing the samples' distribution in a bidimensional plot with the mean native frequency and bandwidth as coordinates and set-

ting two bounds (lower and upper) for each of these two features. Successively, to achieve the best specificity, an additional empirical upper bound was defined for the mean total intensity. Since the empirically binary classifier perfectly overfit the data set, it obviously outperforms the SVM and the decision tree. The scope of this classifier is in fact to show what the best achievable performance for this data set is and not to define a reproducible classification criterion.

9.3 Results

9.3.1 B-Lines' Initial Evaluation

Fig. 9.2 shows how many B-lines were detected for each patient, highlighting the difference between the patients suffering from PF (611 B-lines) and the others (418 B-lines). Indeed, even though the number of B-lines is generally greater in patients suffering from PF, some of these patients (ID 14, 15, and 16) present a low number of artifacts. Fig. 9.3 (top and center) shows how the distributions representing the fibrotic patients are located within a specific range of native frequency and bandwidth, whereas the group of patients affected by

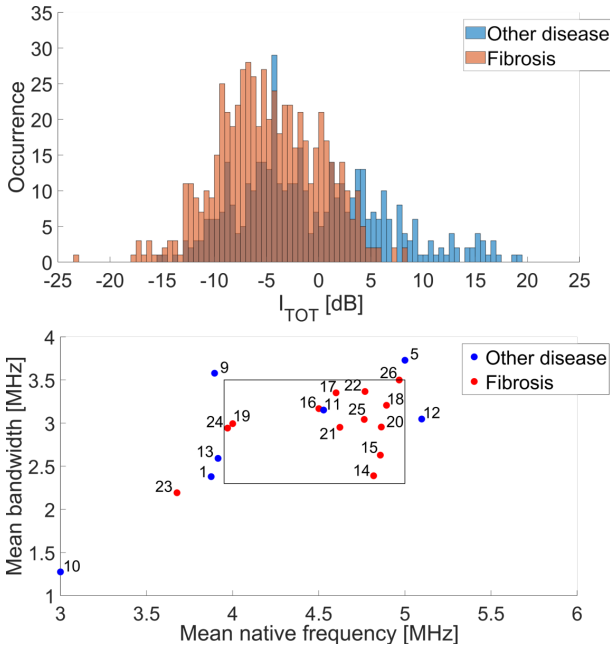

Figure 9.5: Histograms of the total intensity for fibrotic (red bars) and nonfibrotic (blue bars) patients are depicted on the top. The occurrence represents the number of B-lines as a function of the total intensity (x-axis). The 2-D plot (bottom) represents the patients as a function of the mean native frequency and mean bandwidth of the B-lines observed in each patient after the application of a 1.6-dB threshold to the mean total intensity. The nonfibrotic patients are reduced to 7, with a consequent increase of specificity to 92%. The patient IDs are written near the corresponding points. The patients within the black box are classified as fibrotic by the empirically defined binary classifier.

other diseases is more heterogeneous and covers a larger range. To evaluate the possibility to discriminate the fibrotic patients from the others, each patient was represented as a point in a 2-D plane with the mean native frequency and mean bandwidth as coordinates (see Fig. 9.3 at the bottom). Then, a binary classifier was empirically defined so that the patients with the mean native frequency between 3.95 and 5.00 MHz and the mean bandwidth between 2.3 and 3.5 MHz are classified as fibrotic. Despite the simplicity of this classifier, a sensitivity approximately equal to 92% and a specificity about 77% were achieved. To further improve the classification performance, the mean total intensity was evaluated. Indeed, the literature suggests that B-lines probably originate from acoustic traps, whose characteristics depend on the pathologies affecting the lung parenchyma [54]. In particular, in PF, these traps are composed of fibrotic content, whose attenuation is significantly higher than fluids, such as water and blood. Hence, it is expected that B-lines would generally show a lower intensity in fibrotic patients [54], as observable in Fig. 9.4. This concept is further confirmed by the histograms of the total intensity depicted in Fig. 9.5 (top). The mean of the histogram representing the nonfibrotic patients is indeed significantly higher (-0.64 dB) than fibrotic patients (-4.76 dB). As a consequence, to exclude patients with more intense B-lines, a threshold empirically fixed at 1.6 dB was applied to the mean total intensity. The introduction of this new discrimination criterion allowed the correct classification of six nonfibrotic patients, two of whom were previously misclassified. Therefore, as observable in Fig. 9.5 (bottom), thanks to the mean total intensity evaluation, the specificity increased from about 77% to approximately 92%.

9.3.2 Classifiers' Performance

Fig. 9.6 shows the performance of the selected classifiers (SVM and decision tree) in discriminating fibrotic patients from the other group. The best accuracy was obtained when the decision tree was used. In particular, the maximum accuracy (76.9%) was achieved by exploiting the mean native frequency (singular feature) or the mean native frequency together with the bandwidth (multiple features) as parameters. The SVM's accuracy trend is similar, even though the achieved accuracy (maximum value equal to 74% when only the mean native frequency was used as feature) is lower than for the decision tree. The performance of both the decision tree and the SVM decreases when all the features were exploited (maximum accuracy about 72.5% and 68.3%, respectively). This accuracy decrement is likely due to the higher complexity of the model, whose discrimination power hardly increases with such a low number of samples. The decision tree's sensitivity is generally higher than the SVM (maximum sensitivity is, respectively, about 75.8% with the mean native frequency and mean total intensity as features and 71.5% with only the mean native frequency as feature), whereas its maximum specificity is lower than the SVM (respectively, about 86.64% and 92.24% with the mean total intensity as feature).

9.3.3 Features' Statistical Analysis

The p-values obtained by performing the ANOVA test to compare the overall difference between the fibrotic and nonfibrotic groups are 0.0236, 0.3190, and 0.0056 for the mean

native frequency, mean bandwidth, and mean total intensity, respectively.

9.4 Conclusion and Discussion

Due to its portability, cost-effectiveness, and safety, lung ultrasonography represents nowadays a key instrument to provide clinicians with important information on the state of the lung surface. The aforementioned characteristics make LUS extremely useful especially in emergency and critical care settings [24], such as in the monitoring of COVID-19

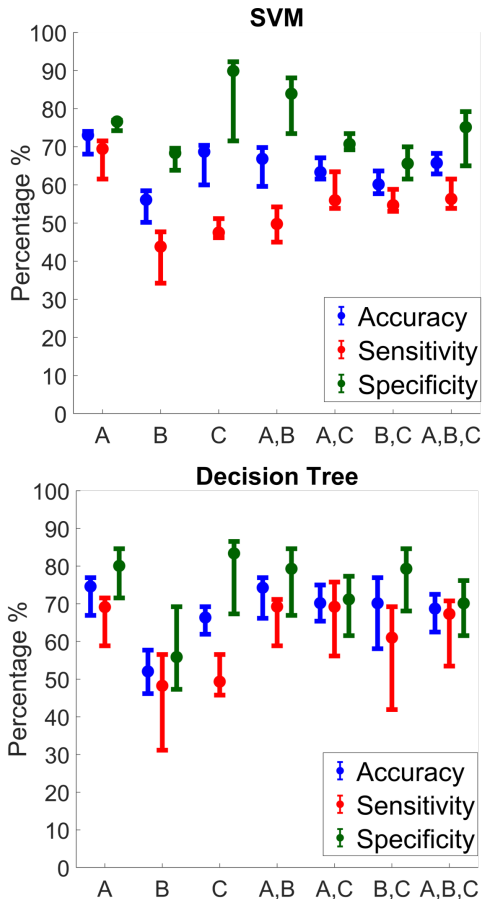

Figure 9.6: Performance of the selected SVM and decision tree is shown. The accuracy, sensitivity with respect to the fibrotic group, and specificity with respect to the fibrotic group are represented as error bars in blue, red, and green, respectively. The upper and lower values of each error bar represent the maximum and minimum values of accuracy, sensitivity, and specificity obtained by varying k parameter of k-fold cross validation. The points of the error bars represent the average value of the performance obtained by averaging the performance values obtained with k-fold cross validation, with k ranging from 2 to 26. The x-axis represents the features exploited in the selected model, where A is the mean native frequency, B is the mean bandwidth, and C is the mean total intensity. When multiple features were exploited, a comma is used in the x-axis to separate them (e.g., A, B).

pneumonia [1, 32]. The potentials of LUS in supporting clinicians during the diagnostic process have been studied since the 1990s, when, for the first time, the vertical artifacts known as B-lines were observed in patients with alveolar interstitial syndrome [17].

B-lines appear generally different in the presence of different pathological conditions of the lung [24]. For example, the possibility to qualitatively differentiate cardiogenic from pneumogenic interstitial syndrome is well recognized in the literature [24]. The appearance of B-lines is indeed different in these two groups of pathologies, likely due to different shape and size of the acoustic traps that form on the diseased lung surface [54].

The characteristics of acoustic traps could be detected by analyzing the frequency spectrum of B-lines [101], which were already proved to be frequency dependent [55]. Of particular interest is the discrimination of patients affected by PF, which is characterized by more attenuated and diffused B-lines, probably due to the heterogeneity and composition of acoustic traps [54].

In this article, we have presented a study investigating the potentials of a B-line frequency characterization method with respect to the differentiation of patients affected by PF. We exploited the multi-frequency approach introduced in [55], which provided images of the same region acquired with four different center frequencies. The raw RF data acquired from 26 patients, half of whom affected by PF, were then processed and the total intensity parameter was extracted to quantitatively evaluate the B-line intensity. To characterize all the 1029 B-lines, the total intensity was evaluated as a function of frequency, hence determining the native frequency and the bandwidth of each B-line.

The results show differences between the vertical artifacts observed in fibrotic patients with respect to patients affected by other diseases (Figs. 7.3 and 7.4). Indeed, while fibrotic patients are more homogeneously distributed within a specific range of native frequencies and bandwidths (native frequency near the average of the investigated frequency range, i.e., 4.5 MHz, and bandwidth within 2.3 and 3.5 MHz), the other group is characterized by a more heterogeneous distribution covering a wider range (Fig. 9.3). This is consistent with the pathological conditions affecting the two groups of patients. The nonfibrotic patients are in fact affected by different diseases (e.g., pneumonia, COPD, and pulmonary hypertension), whereas the fibrotic patients are specifically characterized by their histopathological pattern, which is almost unique. The heterogeneity of the nonfibrotic group is further confirmed by the total intensity histogram (Fig. 9.5 at the top), whose standard deviation is relatively higher (6.9 dB) compared to the other group (4.9 dB). Moreover, the mean total intensity of all the B-lines representing fibrotic patients is significantly lower (-4.76 dB) than the mean value of the nonfibrotic group (-0.64 dB). This difference is likely associated with the physical composition of acoustic traps, which, in several diseases different from PF, are generally consisted of less attenuating media (e.g., water and blood) than fibrotic tissue [50, 140]. As further confirmation, the introduction of a threshold to the mean total intensity allowed the specificity of the empirically defined binary classifier to be improved up to 92% (+15% compared to the binary classification that considers as discriminating parameters only the mean native frequency and bandwidth).

The discrimination power of the three features (native frequency, bandwidth, and total intensity) is reported by the statistical analysis. When investigating the overall

differences between the PF and non-PF populations, a p-value equal to 0.0236, 0.3190, and 0.0056 for the mean native frequency, mean bandwidth, and mean total intensity was obtained, respectively. Considering a threshold to p-values at 0.05, both the native frequency and total intensity are significant for discriminating the two groups. Differently, the bandwidth is not. Furthermore, the statistical analysis' results seem to be consistent with the performance achieved by the SVM and decision tree (Fig. 9.6), where the native frequency has the strongest discrimination power, followed by the total intensity and finally the bandwidth.

In conclusion, the quantitative evaluation of B-lines by means of both the multi-frequency analysis and the total intensity is potentially able to discriminate fibrotic patients. These promising results could open to the definition of an ultrasound imaging technique specifically designed for the lungs. However, there exist cases where the artifactual pattern is similar to fibrotic patients and, hence, the analyzed discrimination parameters (total intensity, native frequency, and bandwidth) are not sufficient. As a consequence, further investigation is necessary. Of particular interest could be the analysis of patients with specific diseases, such as cardiogenic pulmonary edema (CPE), that show artifactual and histopathological patterns extremely different from fibrotic patients. Through the evaluation of a larger group of patients affected by PF and another group of patients affected by CPE, the potentiality of the method described in this article could be further explored.

Chapter 10

Conclusion

This Chapter[1] *draws the conclusions of the thesis within the context of the current LUS literature.*

10.1 Final Discussion

LUS represents a relatively novel application of ultrasound technology, which has been increasingly expanding since the 1990s [17]. However, contrary to standard ultrasound imaging, which was developed primarily for imaging non-invasively the anatomy of internal body parts, LUS is mainly based on the visual interpretation of imaging artifacts. Among these, the so-called vertical artifacts are particularly important as they correlate with various pathologies [19–21]. The main limitations associated with this type of pattern analysis remain its subjectivity and limited reproducibility. Moreover, the understanding and exploitation of the mechanisms underlying the genesis of vertical artifacts are just beginning.

To overcome LUS limits and study vertical artifacts as a mean to characterize the lung surface, various experimental studies have been recently performed [60, 66, 67] and mathematical models have been proposed [58, 101]. Specifically, the acoustic trap theory suggests that vertical artifacts originate from multiple reflections of ultrasound waves trapped within channels that can form between alveoli when lung tissue becomes pathological [58, 101]. By exploiting this concept and the dependence of vertical artifacts on imaging frequency [55, 60], Demi *et al.* proposed a frequency characterization of these artifacts to indirectly estimate the dimension of acoustic channels (or traps). In particular, a lower native frequency should indicate a greater trap size [58, 101]. This theory associates the genesis of vertical artifacts with specific resonance phenomena, as also reported in recent *in vitro* studies [60, 67].

To further investigate the possibility to estimate the dimension of these acoustic traps by means of a frequency characterization of vertical artifacts, we performed an *in silico* study investigating the dependence of vertical artifacts on the alveolar diameter and

[1] This Chapter appears in:
[J0] F. Mento, U. Khan, F. Faita, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "State of The Art in Lung Ultrasound, Shifting From Qualitative To Quantitative Analyses," in *Ultrasound in Medicine and Biology*, in press, 2022.

spacing (Chapter 8) [96]. This study, which is the first *in silico* study on LUS, was of fundamental importance as it allowed the reproduction of vertical artifacts in a numerically simulated domain, whose structure would be hardly reproducible *in vitro*. Indeed, in this study we were able to periodically arrange alveoli at specific distances and having specific diameter sizes. The quantitative results showed that, when spacing was set to half of the wavelength, a possible correlation between the artifacts' intensity and the ratio between the wavelength and spacing was found [96]. This strengthened the potential to estimate the dimension of acoustic traps by means of vertical artifacts' spectroscopy.

The dependence of vertical artifacts on different imaging parameters (center frequency, bandwidth, focal point position and the beam's angle of incidence) was further investigated *in vitro*, both qualitatively [66] and quantitatively [60, 67]. In this regard, as novel contributions, we have presented two *in vitro* studies (Chapters 6 [60] and 7 [67]) that revealed how these artifacts significantly depend on the investigated imaging parameters, regardless of the models utilized to generate them. Indeed, even though the experimental models analyzed in Chapter 6 [60] (microbubble phantoms) completely differ from the model utilized in Chapter 7 [67] (thorax phantom), the results consistently showed a strong dependence of vertical artifacts' intensity on center frequency. Moreover, they showed how the focal point position, number of transmitting elements, bandwidth, and beam's angle of incidence have a non-negligible effect on vertical artifacts' intensity, and could hence be seen as confounding factors. In addition, the study presented in Chapter 6 [60] proved how there was no correlation between the vertical artifacts' intensity and the beam size, thus highlighting how changes in the lateral resolution do not affect vertical artifacts' intensity. These findings were fundamental as they strengthened the hypothesis associating the genesis of vertical artifacts with specific resonance phenomena. As a consequence, these studies further highlighted the potential to exploit the frequency spectrum of vertical artifacts to indirectly estimate the dimension of acoustic traps [60, 67]. These studies also provided important information to reduce the effect of confounding factors. As an example, they recommended positioning the focal point at the pleural line depth, as the structures that can be analyzed are indeed those along the lung surface [60, 67]. The findings of these *in vitro* studies are relevant for the development of quantitative LUS clinical approaches as they quantify the impact of potentially confounding factors. Moreover, the results presented in these studies highlighted how a visual interpretation of these artifacts leads to subjective and qualitative analysis, as their appearance strongly depends on several imaging parameters that are seldom considered in the design of clinical studies. As a consequence, the technique often used in clinical practice, i.e., the count of vertical artifacts in the image, should be considered as qualitative and poorly reproducible.

In contrast, the main advantage of quantitative techniques consists of providing a physical measure able to estimate the state of lung surface, avoiding a subjective evaluation based only on visual interpretation of LUS artifacts. However, even though different quantitative techniques have been proposed since 2017 [55–57], they have not been fully validated in large clinical trials, and specific limitations exist for each technique. The technique proposed by Mohanty *et al.* showed promise in differentiating healthy subjects from patients with lung diseases, but presents a practical limitation related to the impact of the intercostal layer thickness [56]. It has so far been tested only with the probe di-

rectly in contact with the lung. Moreover, it was applied only on rat lungs, which present different characteristics compared to humans [56]. On the other hand, Zhang *et al.* tested LUSWE in human studies and obtained encouraging results in terms of differentiation between healthy subjects and patients affected by different diseases [57, 61, 99, 100]. However, as suggested by an *in vitro* study [92], the positioning of the shaker could lead to strong variability in the findings and, thus, to less reliable results. The multifrequency technique proposed by Demi *et al.* [55] yielded promising results in differentiating patients with PF from patients with different lung diseases [59]. Indeed, consistently with the findings obtained on the *in vitro* and *in silico* studies presented in Chapters 6 [60], 7 [67] and 8 [96], the multifrequency characterization of vertical artifacts proved to be potentially effective also *in vivo* to derive a differential diagnosis (Chapter 9 [59]). Specifically, in this study we showed how the characteristic acoustic traps formed in patients with PF seemed to generate vertical artifacts having native frequencies at specific portions of the spectrum [59]. Nevertheless, only results on a relatively small cohort of patients (26) have been obtained so far.

At present, semi-quantitative techniques still represent the main available strategy to exploit LUS to assess the state of the lung surface. Nevertheless, the use of these techniques should be guided by proper definition and standardization of acquisition protocols. Specifically, given the dependence of vertical artifacts on imaging parameters [60, 66, 67], standardization of the imaging protocols represents a key methodology to reduce confounding factors. Unfortunately, many protocols are heterogeneously defined, often lacking details on the imaging settings [27, 35, 42]. Also the scoring systems, as well as the amount and location of the scanning areas, are often arbitrarily defined [25]. Differently, to develop and validate proper imaging protocols and scoring systems to be used depending on the disease being investigated, it is of fundamental importance to compare the performance of different protocols following an evidence-based approach [28, 29]. To this end, we performed two studies analyzing the impact of both location and amount of scanning areas on the evaluation of COVID-19 [28] (Chapter 4) and post-COVID-19 patients [29] (Chapter 5). Specifically, we exploited a dataset including LUS examinations performed on 220 patients (100 COVID-19 positive and 120 post-COVID-19) by utilizing a 14-areas standardized and clinically validated imaging protocol [1, 38]. The results showed how the scanning areas could be reduced to 10 without significantly impacting the final evaluation of the patient, as long as the basal posterior areas are scanned [29]. These important findings could allow the development of specific LUS acquisition protocols, characterized by an optimal trade-off between time-efficiency and accuracy.

An additional problem related to semi-quantitative approaches is the subjectivity of the analysis. Indeed, even though scoring systems are used, they are based only on the coding of visually interpreted LUS patterns into scores. Hence, the operator dependence is impossible to overcome. To this end, the use of AI to automatically score LUS data could be instrumental in reducing the subjectivity in the evaluation of LUS patterns. For this reason, in Chapter 3 we proposed a technique to automatically classify LUS videos by means of an aggregation strategy [45] based on a deep learning (DL) frame-based classification [2]. The development of this technique was fundamental as it allowed the comparison between clinicians' and AI's evaluations at video and prognostic levels.

Specifically, the results at prognostic level showed the potential to exploit AI algorithm in the automatic stratification of COVID-19 positive patients [45] and post-COVID-19 patients [29]. Nevertheless, given the strong subjectivity of the task [3, 37], it is not reliable to expect levels of agreement between AI and human operators at video level around 90-100% [29, 45]. Indeed, the use of AI algorithms could lead to more reproducible analyses but cannot completely avoid subjectivity, as AI training remains based on the subjective labeling performed by clinicians.

Another problem that emerged during the COVID-19 pandemic was the extensive development of AI solutions regardless of domain knowledge on LUS. As an example, AI systems have been proposed for the diagnosis of COVID-19 based on the simple evaluation of LUS videos or frames [144–146]. This application of AI is fallacious as it is not possible to diagnose COVID-19 based on LUS. Indeed, positive patients may still not present any alterations along the lung surface. Moreover, LUS imaging patterns are strongly unspecific [42]. As an example, different LUS artifacts (e.g., vertical artifacts and *"white lung"*) or anatomical findings (e.g, consolidations) that were observed in COVID-19 patients could be observed in many other diseases (e.g., CPE, PF, and ARDS).

In conclusion, quantitative approaches represent the future of LUS, as they could provide a physical metric able to characterize the lung surface by applying an acquisition technique specifically designed for the lung. This was demonstrated in Chapter 9, where a novel multifrequency quantitative LUS strategy was shown to be effective in discriminating patients with PF from patients with other lung diseases [59]. Nevertheless, to further develop these techniques, the genesis of vertical artifacts needs to be more deeply investigated and understood by means of controlled *in vitro* and *in silico* studies, as done in the studies presented in Chapters 6, 7, and 8. In the meantime, semi-quantitative approaches based on image analysis techniques should be exploited to estimate the state of lungs by detecting and recognizing specific LUS patterns that do signal different levels of aeration. However, to reduce the impact of confounding factors, standardization of the imaging protocols and scoring systems is essential. Moreover, it is fundamental to compare the performance of different protocols to properly define an acquisition protocol able to optimize both execution time and accuracy of the examination (Chapter 4 [28] and Chapter 5 [29]). In this context, consistently with what was done in Chapter 3 [45], AI algorithms could be used to guide the analysis of LUS data.

List of Publications

INTERNATIONAL JOURNALS

[J0] **F. Mento**, U. Khan, F. Faita, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "State of The Art in Lung Ultrasound, Shifting From Qualitative To Quantitative Analyses," in *Ultrasound in Medicine and Biology*, in press, 2022 [147].

[J1] **F. Mento**, T. Perrone, A. Fiengo, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, "Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study," in *The Journal of Acoustical Society of America*, vol. 149, no. 5, pp. 3626–3634, May 2021 [45].

[J2] **F. Mento**, T. Perrone, V. N. Macioce, F. Tursi, D. Buonsenso, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, "On the Impact of Different Lung Ultrasound Imaging Protocols in the Evaluation of Patients Affected by Coronavirus Disease 2019: How Many Acquisitions Are Needed?," in *Journal of Ultrasound in Medicine*, vol. 40, no. 10, pp. 2235-2238, October 2021 [28].

[J3] L. Demi, **F. Mento**, A. Di Sabatino, A. Fiengo, U. Sabatini, V. N. Macioce, M. Robol, F. Tursi, C. Sofia, C. Di Cienzo, A. Smargiassi, R. Inchingolo, T. Perrone, "Lung Ultrasound in COVID-19 and Post-COVID-19 Patients, an Evidence-Based Approach," in *Journal of Ultrasound in Medicine*, Online ahead of print, December 2021 [29].

[J4] **F. Mento**, L. Demi, "On the influence of imaging parameters on lung ultrasound B-line artifacts, *in vitro* study," in *The Journal of Acoustical Society of America*, vol. 148, no. 2, pp. 975-983, August 2020 [60].

[J5] **F. Mento**, L. Demi, "Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: An *in vitro* study," in *The Journal of Acoustical Society of America*, vol. 150, no. 6, pp. 4075-4082, December 2021 [67].

[J6] E. Peschiera, **F. Mento**, L. Demi, "Numerical study on lung ultrasound B-line formation as a function of imaging frequency and alveolar geometries," in *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2304–2311, April 2021 [96].

[J7] **F. Mento**, G. Soldati, R. Prediletto, M. Demi, L. Demi, "Quantitative Lung Ultrasound Spectroscopy Applied to the Diagnosis of Pulmonary Fibrosis: The First

Clinical Study," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2265-2273, November 2020 [59].

[J8] 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. V. Sloun, E. Ricci, and L. Demi, "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," in *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2676–2687, 2020 [2].

[J9] O. Frank, N. Schipper, M. Vaturi, G. Soldati, A. Smargiassi, R. Inchingolo, E. Torri, T. Perrone, **F. Mento**, L. Demi, M. Galun, Y. C. Eldar, and S. Bagon, "Integrating Domain Knowledge into Deep Networks for Lung Ultrasound with Applications to COVID-19," in *IEEE Transactions on Medical Imaging*, Online ahead of print, 2021 [47].

[J10] R. Roshankhah, Y. Karbalaeisadegh, G. Hastings, **F. Mento**, G. Soldati, A. Smargiassi, R. Inchingolo, E. Torri, T. Perrone, S. Aylward, L. Demi, and M. Muller, "Investigating training-test data splitting strategies for automated segmentation and scoring of COVID-19 lung ultrasound images," in *The Journal of The Acoustical Society of America*, vol. 150, no. 6, pp. 4118–4127, 2021 [48].

[J11] L. Carrer, E. Donini, D. Marinelli, M. Zanetti, **F. Mento**, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, F. Bovolo, and L. Bruzzone, "Automatic Pleural Line Extraction and COVID-19 Scoring from Lung Ultrasound Data," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2207–2217, 2020 [43].

[J12] 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**, and L. Demi, "Is There a Role for Lung Ultrasound During the COVID-19 Pandemic?" in *Journal of Ultrasound in Medicine*, vol. 39, no. 7, pp. 1459–1462, 2020 [32].

[J13] 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**, and L. Demi, "Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19," in *Journal of Ultrasound in Medicine*, vol. 39, no. 7, pp. 1413–1419, 2020 [1].

[J14] A. Smargiassi, G. Soldati, E. Torri, **F. Mento**, D. Milardi, P. D. Giacomo, G. De Matteis, M. L. Burzo, A. R. Larici, M. Pompili, L. Demi, and R. Inchingolo, "Lung Ultrasound for COVID-19 Patchy Pneumonia: Extended or Limited Evaluations?" in *Journal of Ultrasound in Medicine*, vol. 40, no. 3, pp. 521–528, August 2020 [30].

[J15] **F. Mento**, T. Perrone, A. Fiengo, F. Tursi, V. N. Macioce, A. Smargiassi, R. Inchingolo, and L. Demi, "Limiting the areas inspected by lung ultrasound leads to an underestimation of COVID-19 patients' condition," in *Intensive Care Medicine*, vol. 47, no. 7, pp. 811–812, 2021 [118].

[J16] 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**, and L. Demi, "Time for a new international evidence-based recommendations for point-of-care lung ultrasound," in *Journal of Ultrasound in Medicine*, vol. 40, no. 2, pp. 433–434, August 2021 [31].

[J17] G. Soldati, A. Smargiassi, T. Perrone, E. Torri, **F. Mento**, L. Demi, and R. Inchingolo, "There is a Validated Acquisition Protocol for Lung Ultrasonography in COVID-19 Pneumonia," in *Journal of Ultrasound in Medicine*, vol. 40, no. 12, p. 2783, February 2021 [41].

[J18] G. Soldati, A. Smargiassi, T. Perrone, E. Torri, **F. Mento**, L. Demi, and R. Inchingolo "LUS for COVID-19 Pneumonia: Flexible or Reproducible Approach?" in *Journal of Ultrasound in Medicine*, vol. 41, no. 2, pp. 525–526, April 2021 [40].

[J19] 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**, and L. Demi, "On Lung Ultrasound Patterns Specificity in the Management of COVID-19 Patients," in *Journal of Ultrasound in Medicine*, vol. 39, no. 11, pp. 2283–2284, May 2020 [148].

[J20] U. Khan, **F. Mento**, L. N. Giacomaz, R. Trevisan, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "Deep learning-based classification of reduced lung ultrasound data from COVID-19 patients," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 5, pp. 1661-1669, March 2022 [83].

[J21] N. Fatima, **F. Mento**, A. Zanforlin, A. Smargiassi, E. Torri, T. Perrone, and L. Demi, "Human to AI Interrater Agreement for Lung Ultrasound Scoring in COVID-19 Patients," in *Journal of Ultrasound in Medicine*, Online ahead of print, July 2022 [3].

INTERNATIONAL CONFERENCE PROCEEDINGS

[CP1] **F. Mento**, G. Soldati, R. Prediletto, M. Demi, and L. Demi, "Differentiation of Pulmonary Fibrosis by Means of Quantitative Lung Ultrasound Spectroscopy, First Clinical Study in Humans," in *2020 IEEE International Ultrasonics Symposium (IUS)*, Las Vegas, USA, 2020, pp. 1–4 [149].

[CP2] **F. Mento** and L. Demi, "Effect of Imaging Parameters on the Visualization of Lung Ultrasound B-line Artifacts," in *2020 IEEE International Ultrasonics Symposium (IUS)*, Las Vegas, USA, 2020, pp. 1–4 [150].

[CP3] D. Yaron, D. Keidar, E. Goldstein, Y. Shachar, A. Blass, O. Frank, N. Schipper, N. Shabshin, A. Grubstein, D. Suhami, N. R. Bogot, C. S. Weiss, E. Sela, A. A. Dror, M. Vaturi, **F. Mento**, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, T. Perrone, L. Demi, M. Galun, S. Bagon, Y. M. Elyada, and Y. C. Eldar, "Point of Care Image Analysis for COVID-19," in *ICASSP 2021 - 2021 IEEE International Conference*

on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Canada, 2021, pp. 8153–8157 [151].

[CP4] **F. Mento**, T. Perrone, A. Fiengo, V. N. Macioce, F. Tursi, A. Smargiassi, R. Inchingolo, and L. Demi, "A Multicenter Study Assessing Artificial Intelligence Capability in Scoring Lung Ultrasound Videos of COVID-19 Patients," in *2021 IEEE International Ultrasonics Symposium (IUS)*, Xi'an, China, 2021, pp. 1–3 [152].

[CP5] **F. Mento** and L. Demi, "Impact of Frequency, Bandwidth, Focus, and Angle of Incidence on Lung Ultrasound Vertical Artifacts' Intensity, in-vitro Study," in *2021 IEEE International Ultrasonics Symposium (IUS)*, Xi'an, China, 2021, pp. 1–3 [153].

[CP6] U. Khan, **F. Mento**, L. N. Giacomaz, R. Trevisan, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "Impact of pixel, intensity, and temporal resolution on automatic scoring of LUS from Coronavirus disease 2019 patients," in *Proceedings of Meetings on Acoustics*, in press, 2022 [154].

[CP7] L. L. Custode, **F. Mento**, S. Afrakhteh, F. Tursi, A. Smargiassi, R. Inchingolo, T. Perrone, G. Iacca, and L. Demi, "Neuro-symbolic interpretable AI for automatic COVID-19 patient-stratification based on standardised lung ultrasound data," Proceedings of Meetings on Acoustics, in press, 2022 [155].

INTERNATIONAL CONFERENCE ABSTRACTS

[CA1] **F. Mento**, G. Soldati, R. Prediletto, M. Demi, and L. Demi, "The impact of B-lines' frequency characterization on lung ultrasound imaging, *in vitro*- and *in vivo* study," in *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2692, 2020. [Online]. Available: `https://doi.org/10.1121/1.5147454`

[CA2] R. Roshankhah, Y. Karbalaeisadegh, H. Greer, **F. Mento**, G. Soldati, A. Smargiassi, R. Inchingolo, E. Torri, S. Aylward, L. Demi, and M. Muller, "Automated segmentation and scoring of lung ultrasound images of COVID-19 patients," in *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2735, 2020. [Online]. Available: `https://doi.org/10.1121/1.5147599`

[CA3] A. Smargiassi, G. Soldati, T. Perrone, E. Torri, **F. Mento**, D. Milardi, P. Del Giacomo, G. De Matteis, M. L. Burzo, A. R. Larici, M. Pompili, L. Demi, and R. Inchingolo, "Lung ultrasound and high-resolution CT-scan of the chest for COVID-19 pneumonia," in *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2691, 2020. [Online]. Available: `https://doi.org/10.1121/1.5147452`

[CA4] S. Bagon, M. Galun, O. Frank, N. Schipper, M. Vaturi, G. Zalcberg, G. Soldati, A. Smargiassi, R. Inchingolo, E. Torri, T. Perrone, **F. Mento**, L. Demi, and Y. Eldar, "Assessment of COVID-19 in lung ultrasound by combining anatomy and sonographic artifacts using deep learning," in *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2736, 2020. [Online]. Available: `https://doi.org/10.1121/1.5147600`

[CA5] E. Peschiera, T. Rigolin, **F. Mento**, and L. Demi, "Ultrasound waves propagation in aerated inhomogeneous media," in *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2737, 2020. [Online]. Available: https://doi.org/10.1121/1.5147604

[CA6] L. Demi, **F. Mento**, T. Perrone, A. Fiengo, A. Smargiassi, R. Inchingolo, and G. Soldati, "Agreement between expert sonographers and artificial intelligence in the evaluation of lung ultrasound data acquired from COVID-19 patients," ERJ Open Research, vol. 7, no. suppl 6, 2021. [Online]. Available: https://doi.org/10.1183/23120541.LSC-2021.61

[CA7] **F. Mento**, and L. Demi, "Investigating the link between intensity of lung ultrasound vertical artifacts and penetration depth of ultrasound waves, *in silico* study," in *The Journal of the Acoustical Society of America*, vol. 151, no. 4, pp. A76, 2022. [Online]. Available: https://doi.org/10.1121/10.0010706

[CA8] L. Demi, **F. Mento**, A. Di Sabatino, A. Fiengo, U. Sabatini, V. N. Macioce, M. Robol, F. Tursi, C. Sofia, C. Di Cienzo, A. Smargiassi, R. Inchingolo, and T. Perrone, "A standardised and evidenced-based lung ultrasound acquisition protocol and scoring system for the monitoring and stratification of COVID-19 and post-COVID-19 patients," in *The Journal of the Acoustical Society of America*, vol. 151, no. 4, pp. A112, 2022. [Online]. Available: https://doi.org/10.1121/10.0010818

[CA9] U. Khan, **F. Mento**, L. Nicolussi Giacomaz, R. Trevisan, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "On the impact of pixel resolution on automated scoring of lung ultrasound images from coronavirus disease 2019 patients," in *The Journal of the Acoustical Society of America*, vol. 151, no. 4, pp. A112, 2022. [Online]. Available: https://doi.org/10.1121/10.0010819

[CA10] L. L. Custode, **F. Mento**, S. Afrakhteh, F. Tursi, A. Smargiassi, R. Inchingolo, T. Perrone, L. Demi, and G. Iacca, "Neuro-symbolic interpretable AI for automatic COVID-19 patient-stratification based on standardised lung ultrasound data," in *The Journal of the Acoustical Society of America*, vol. 151, no. 4, pp. A112, 2022. [Online]. Available: https://doi.org/10.1121/10.0010820

[CA11] **F. Mento**, L. Demi, A. Di Sabatino, A. Fiengo, U. Sabatini, V. N. Macioce, M. Robol, F. Tursi, C. Sofia, C. Di Cienzo, A. Smargiassi, R. Inchingolo, and T. Perrone, "Multicenter Study on Lung Ultrasound COVID-19 and Post-COVID-19 Patients," in *Journal of Ultrasound in Medicine*, vol. 41, no. S1, pp. S129, 2022. [Online]. Available: https://doi.org/10.1002/jum.16028

[CA12] **F. Mento**, U. Khan, L. Nicolussi Giacomaz, R. Trevisan, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "On the Impact of Pixel Resolution on Automated Scoring of Lung Ultrasound Images From COVID-19 Patients," in *Journal of Ultrasound in Medicine*, vol. 41, no. S1, pp. S130, 2022. [Online]. Available: https://doi.org/10.1002/jum.16028

Bibliography

[1] 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, and L. Demi, "Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19," *Journal of Ultrasound in Medicine*, vol. 39, no. 7, pp. 1413–1419, 2020. [Online]. Available: https://doi.org/10.1002/jum.15285

[2] 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. V. Sloun, E. Ricci, and L. Demi, "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2676–2687, 2020. [Online]. Available: https://doi.org/10.1109/TMI.2020.2994459

[3] N. Fatima, F. Mento, A. Zanforlin, A. Smargiassi, E. Torri, T. Perrone, and L. Demi, "Human to AI Interrater Agreement for Lung Ultrasound Scoring in COVID-19 Patients," *Journal of Ultrasound in Medicine*, July 2022. [Online]. Available: https://doi.org/10.1002/jum.16052

[4] F. Dunn and W. J. Fry, "Ultrasonic absorption and reflection by lung tissue," *Physics in Medicine and Biology*, vol. 5, no. 4, pp. 401–410, 1961. [Online]. Available: https://doi.org/10.1088/0031-9155/5/4/302

[5] T. J. Bauld and H. P. Schwan, "Attenuation and reflection of ultrasound in canine lung tissue," *Journal of the Acoustical Society of America*, vol. 56, no. 5, pp. 1630–1637, 1974. [Online]. Available: https://doi.org/10.1121/1.1903488

[6] F. Dunn, "Attenuation and speed of ultrasound in lung," *Journal of the Acoustical Society of America*, vol. 56, no. 5, pp. 1638–1639, 1974. [Online]. Available: https://doi.org/10.1121/1.1903489

[7] ——, "Attenuation and speed of ultrasound in lung: Dependence upon frequency and inflation," *Journal of the Acoustical Society of America*, vol. 80, no. 4, pp. 1248–1250, 1986. [Online]. Available: https://doi.org/10.1121/1.393818

[8] P. C. Pedersen and H. S. Ozcan, "Ultrasound properties of lung tissue and their measurements," *Ultrasound in Medicine and Biology*, vol. 12, no. 6, pp. 483–499, 1986. [Online]. Available: https://doi.org/10.1016/0301-5629(86)90220-6

[9] R. T. Towa, R. J. Miller, L. A. Frizzell, J. F. Zachary, and W. D. O'Brien Jr., "Attenuation coefficient and propagation speed estimates of rat and pig intercostal tissue as a function of temperature," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, no. 10, pp. 1411–1420, 2002. [Online]. Available: https://doi.org/10.1109/TUFFC.2002.1041082

[10] M. L. Oelze, R. J. Miller, J. P. Blue Jr., J. F. Zachary, and W. D. O'Brien Jr., "Estimation of the acoustic impedance of lung versus level of inflation for different species and ages of animals," *Journal of the Acoustical Society of America*, vol. 124, no. 4, pp. 2340–2352, 2008. [Online]. Available: https://doi.org/10.1121/1.2973186

[11] K. B. Sagar, T. L. Rhyne, G. S. Myers, and R. S. Lees, "Characterization of Normal and Abnormal Pulmonary Surface by Reflected Ultrasound," *Chest*, vol. 74, no. 1, pp. 29–33, 1978. [Online]. Available: https://doi.org/10.1378/chest.74.1.29

[12] C. R. Joyner Jr., L. D. Miller, S. J. Dudrick, D. J. Eskin, and P. Bloom, "Reflected ultrasound in the study of diseases of the chest." *Transactions of the American Clinical and Climatological Association*, vol. 78, pp. 28–37, 1967. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2441130/

[13] L. D. Miller, C. R. Joyner Jr., S. J. Dudrick, and D. J. Eskin, "Clinical use of ultrasound in the early diagnosis of pulmonary embolism." *Annals of surgery*, vol. 166, no. 3, pp. 381–393, 1967. [Online]. Available: https://dx.doi.org/10.1097%2F00000658-196709000-00006

[14] M. C. Ziskin, D. I. Thickman, N. J. Goldenberg, M. S. Lapayowker, and J. M. Becker, "The comet tail artifact," *Journal of Ultrasound in Medicine*, vol. 1, no. 1, pp. 1–7, 1982. [Online]. Available: https://doi.org/10.7863/jum.1982.1.1.1

[15] D. I. Thickman, M. C. Ziskin, N. Jacobs Goldenberg, and B. E. Linder, "Clinical manifestations of the comet tail artifact," *Journal of Ultrasound in Medicine*, vol. 2, no. 5, pp. 225–230, 1983. [Online]. Available: https://doi.org/10.7863/jum.1983.2.5.225

[16] L. Avruch and P. L. Cooperberg, "The ring-down artifact," *Journal of Ultrasound in Medicine*, 1985. [Online]. Available: https://doi.org/10.7863/jum.1985.4.1.21

[17] D. Lichtenstein, G. Mézière, P. Biderman, A. Gepner, and O. Barré, "The comet-tail artifact: An ultrasound sign of alveolar-interstitial syndrome," *American Journal of Respiratory and Critical Care Medicine*, vol. 156, no. 5, pp. 1640–1646, 1997. [Online]. Available: https://doi.org/10.1164/ajrccm.156.5.96-07096

[18] G. Mathis and K. Dirschmid, "Pulmonary infarction: sonographic appearance with pathologic correlation," *European Journal of Radiology*, vol. 17, no. 3, pp. 170–174, 1993. [Online]. Available: https://doi.org/10.1016/0720-048x(93)90098-8

[19] Z. Jambrik, S. Monti, V. Coppola, E. Agricola, G. Mottola, M. Miniati, and E. Picano, "Usefulness of ultrasound lung comets as a nonradiologic sign of extravascular lung water," *American Journal of Cardiology*, vol. 93, no. 10, pp. 1265–1270, 2004. [Online]. Available: https://doi.org/10.1016/j.amjcard.2004.02.012

[20] E. Picano, F. Frassi, E. Agricola, S. Gligorova, L. Gargani, and G. Mottola, "Ultrasound lung comets: A clinically useful sign of extravascular lung water," *Journal of the American Society of Echocardiography*, vol. 19, no. 3, pp. 356–363, 2006. [Online]. Available: https://doi.org/10.1016/j.echo.2005.05.019

[21] R. Copetti, G. Soldati, and P. Copetti, "Chest sonography: a useful tool to differentiate acute cardiogenic pulmonary edema from acute respiratory distress syndrome," *Cardiovascular Ultrasound*, vol. 6, no. 1, p. 16, 2008. [Online]. Available: https://doi.org/10.1186/1476-7120-6-16

[22] D. A. Lichtenstein, N. Lascols, G. Mezière, and A. Gepner, "Ultrasound diagnosis of alveolar consolidation in the critically ill," *Intensive Care Medicine*, vol. 30, no. 2, pp. 276–281, 2004. [Online]. Available: https://doi.org/10.1007/s00134-003-2075-6

[23] G. Volpicelli, M. Elbarbary, M. Blaivas, D. A. Lichtenstein, G. Mathis, A. W. Kirkpatrick, L. Melniker, L. Gargani, V. E. Noble, G. Via, A. Dean, J. W. Tsung, G. Soldati, R. Copetti, B. Bouhemad, A. Reissig, E. Agricola, J.-J. Rouby, C. Arbelot, A. Liteplo, A. Sargsyan, F. Silva, R. Hoppmann, R. Breitkreutz, A. Seibel, L. Neri, E. Storti, and T. Petrovic, "International evidence-based recommendations for point-of-care lung ultrasound." *Intensive care medicine*, vol. 38, no. 4, pp. 577–591, April 2012. [Online]. Available: https://doi.org/10.1007/s00134-012-2513-4

[24] G. Soldati, M. Demi, A. Smargiassi, R. Inchingolo, and L. Demi, "The role of ultrasound lung artifacts in the diagnosis of respiratory diseases," *Expert Review of Respiratory Medicine*, vol. 13, no. 2, pp. 163–172, February 2019. [Online]. Available: https://doi.org/10.1080/17476348.2019.1565997

[25] M. Allinovi, A. Parise, M. Giacalone, A. Amerio, M. Delsante, A. Odone, A. Franci, F. Gigliotti, S. Amadasi, D. Delmonte, N. Parri, and A. Mangia, "Lung Ultrasound May Support Diagnosis and Monitoring of COVID-19 Pneumonia," *Ultrasound in Medicine and Biology*, vol. 46, no. 11, pp. 2908–2917, September 2020. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2020.07.018

[26] N. Bonadia, A. Carnicelli, A. Piano, D. Buonsenso, E. Gilardi, C. Kadhim, E. Torelli, M. Petrucci, L. Di Maurizio, D. G. Biasucci, M. Fuorlo, E. Forte, R. Zaccaria, and F. Franceschi, "Lung Ultrasound Findings Are Associated with Mortality and Need for Intensive Care Admission in COVID-19 Patients Evaluated in the Emergency Department," *Ultrasound in Medicine and Biology*, vol. 46, no. 11, pp. 2927–2937, 2020. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2020.07.005

[27] A. Dargent, E. Chatelain, L. Kreitmann, J.-P. Quenot, M. Cour, L. Argaud, M. Antoine, T. Baudry, P.-J. Bertrand, J. Bougnaud, F. Charbonnieras, M. Damien, C. Delsol, P. Gimet, A. Moutou, E. Nourry, A. Pinede, Z. Lougnon, H. Roccia, M. Romain, M. Simon, K. Srage, and M. Venet, "Lung ultrasound score to monitor COVID-19 pneumonia progression in patients with ARDS," *PLoS ONE*, vol. 15, no. 7

July, 2020. [Online]. Available: https://dx.doi.org/10.1371%2Fjournal.pone.0236312

[28] F. Mento, T. Perrone, V. N. Macioce, F. Tursi, D. Buonsenso, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, and L. Demi, "On the Impact of Different Lung Ultrasound Imaging Protocols in the Evaluation of Patients Affected by Coronavirus Disease 2019: How Many Acquisitions Are Needed?" *Journal of Ultrasound in Medicine*, vol. 40, no. 10, pp. 2235–2238, October 2021. [Online]. Available: https://doi.org/10.1002/jum.15580

[29] L. Demi, F. Mento, A. Di Sabatino, A. Fiengo, U. Sabatini, V. N. Macioce, M. Robol, F. Tursi, C. Sofia, C. Di Cienzo, A. Smargiassi, R. Inchingolo, and T. Perrone, "Lung Ultrasound in COVID-19 and Post-COVID-19 Patients, an Evidence-Based Approach," *Journal of Ultrasound in Medicine*, December 2021. [Online]. Available: https://doi.org/10.1002/jum.15902

[30] A. Smargiassi, G. Soldati, E. Torri, F. Mento, D. Milardi, P. D. Giacomo, G. De Matteis, M. L. Burzo, A. R. Larici, M. Pompili, L. Demi, and R. Inchingolo, "Lung Ultrasound for COVID-19 Patchy Pneumonia: Extended or Limited Evaluations?" *Journal of Ultrasound in Medicine*, vol. 40, no. 3, pp. 521–528, August 2020. [Online]. Available: https://doi.org/10.1002/jum.15428

[31] 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, and L. Demi, "Time for a new international evidence-based recommendations for point-of-care lung ultrasound," *Journal of Ultrasound in Medicine*, vol. 40, no. 2, pp. 433–434, August 2021. [Online]. Available: https://doi.org/10.1002/jum.15412

[32] ——, "Is There a Role for Lung Ultrasound During the COVID-19 Pandemic?" *Journal of Ultrasound in Medicine*, vol. 39, no. 7, pp. 1459–1462, 2020. [Online]. Available: https://doi.org/10.1002/jum.15284

[33] G. Soldati, G. Giannasi, A. Smargiassi, R. Inchingolo, and L. Demi, "Contrast-Enhanced Ultrasound in Patients With COVID-19," *Journal of Ultrasound in Medicine*, vol. 39, no. 12, pp. 2483–2489, 2020. [Online]. Available: https://doi.org/10.1002/jum.15338

[34] L. R. Sultan and C. M. Sehgal, "A Review of Early Experience in Lung Ultrasound in the Diagnosis and Management of COVID-19," *Ultrasound in Medicine and Biology*, vol. 46, no. 9, pp. 2530–2545, 2020. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2020.05.012

[35] L. Zhao, K. Yu, Q. Zhao, R. Tian, H. Xie, L. Xie, P. Deng, G. Xie, A. Bao, and J. Du, "Lung Ultrasound Score in Evaluating the Severity of Coronavirus Disease 2019 (COVID-19) Pneumonia," *Ultrasound in Medicine and Biology*, vol. 46, no. 11, pp. 2938–2944, November 2020. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2020.07.024

[36] L. Demi, F. Mento, T. Perrone, A. Fiengo, A. Smargiassi, R. Inchingolo, and G. Soldati, "Agreement between expert sonographers and artificial intelligence in the evaluation of lung ultrasound data acquired from COVID-19 patients," *ERJ Open Research*, vol. 7, no. suppl 6, p. 61, 2021. [Online]. Available: https://doi.org/10.1183/23120541.LSC-2021.61

[37] M. H. Lerchbaumer, J. H. Lauryn, U. Bachmann, P. Enghard, T. Fischer, J. Grune, N. Hegemann, D. Khadzhynov, J. M. Kruse, L. J. Lehner, T. Lindner, T. Oezkan, D. Zickler, W. M. Kuebler, B. Hamm, K.-U. Eckardt, and F. Muench, "Point-of-care lung ultrasound in COVID-19 patients: inter- and intra-observer agreement in a prospective observational study," *Scientific Reports*, vol. 11, no. 1, 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-90153-2

[38] T. Perrone, G. Soldati, L. Padovini, A. Fiengo, G. Lettieri, U. Sabatini, G. Gori, F. Lepore, M. Garolfi, I. Palumbo, R. Inchingolo, A. Smargiassi, L. Demi, E. E. Mossolani, F. Tursi, C. Klersy, and A. Di Sabatino, "A New Lung Ultrasound Protocol Able to Predict Worsening in Patients Affected by Severe Acute Respiratory Syndrome Coronavirus 2 Pneumonia," *Journal of Ultrasound in Medicine*, vol. 40, no. 8, pp. 1627–1635, 2021. [Online]. Available: https://doi.org/10.1002/jum.15548

[39] F. M. Russell, R. R. Ehrman, A. Barton, E. Sarmiento, J. E. Ottenhoff, and B. K. Nti, "B-line quantification: comparing learners novice to lung ultrasound assisted by machine artificial intelligence technology to expert review," *Ultrasound Journal*, vol. 13, no. 1, 2021. [Online]. Available: https://doi.org/10.1186/s13089-021-00234-6

[40] G. Soldati, A. Smargiassi, T. Perrone, E. Torri, F. Mento, L. Demi, and R. Inchingolo, "LUS for COVID-19 Pneumonia: Flexible or Reproducible Approach?" *Journal of Ultrasound in Medicine*, vol. 41, no. 2, pp. 525–526, April 2021. [Online]. Available: https://doi.org/10.1002/jum.15726

[41] ——, "There is a Validated Acquisition Protocol for Lung Ultrasonography in COVID-19 Pneumonia," *Journal of Ultrasound in Medicine*, vol. 40, no. 12, p. 2783, February 2021. [Online]. Available: https://doi.org/10.1002/jum.15649

[42] L. Demi, "Lung ultrasound: The future ahead and the lessons learned from COVID-19," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2146–2150, October 2020. [Online]. Available: https://doi.org/10.1121/10.0002183

[43] L. Carrer, E. Donini, D. Marinelli, M. Zanetti, F. Mento, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, L. Demi, F. Bovolo, and L. Bruzzone, "Automatic Pleural Line Extraction and COVID-19 Scoring from Lung Ultrasound Data," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2207–2217, 2020. [Online]. Available: https://doi.org/10.1109/tuffc.2020.3005512

[44] J. Chen, C. He, J. Yin, J. Li, X. Duan, Y. Cao, L. Sun, M. Hu, W. Li, and Q. Li, "Quantitative Analysis and Automated Lung Ultrasound Scoring for Evaluating COVID-19 Pneumonia with Neural Networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 7, pp. 2507–2515, 2021. [Online]. Available: https://doi.org/10.1109/tuffc.2021.3070696

[45] F. Mento, T. Perrone, A. Fiengo, A. Smargiassi, R. Inchingolo, G. Soldati, and L. Demi, "Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study," *The Journal of the Acoustical Society of America*, vol. 149, no. 5, pp. 3626–3634, 2021. [Online]. Available: https://doi.org/10.1121/10.0004855

[46] W. Xue, C. Cao, J. Liu, Y. Duan, H. Cao, J. Wang, X. Tao, Z. Chen, M. Wu, J. Zhang, H. Sun, Y. Jin, X. Yang, R. Huang, F. Xiang, Y. Song, M. You, W. Zhang, L. Jiang, Z. Zhang, S. Kong, Y. Tian, L. Zhang, D. Ni, and M. Xie, "Modality alignment contrastive learning for severity assessment of COVID-19 from lung ultrasound and clinical information," *Medical Image Analysis*, vol. 69, 2021. [Online]. Available: https://doi.org/10.1016/j.media.2021.101975

[47] O. Frank, N. Schipper, M. Vaturi, G. Soldati, A. Smargiassi, R. Inchingolo, E. Torri, T. Perrone, F. Mento, L. Demi, M. Galun, Y. C. Eldar, and S. Bagon, "Integrating Domain Knowledge into Deep Networks for Lung Ultrasound with Applications to COVID-19," *IEEE Transactions on Medical Imaging*, 2021. [Online]. Available: https://doi.org/10.1109/tmi.2021.3117246

[48] R. Roshankhah, Y. Karbalaeisadegh, G. Hastings, F. Mento, G. Soldati, A. Smargiassi, R. Inchingolo, E. Torri, T. Perrone, S. Aylward, L. Demi, and M. Muller, "Investigating training-test data splitting strategies for automated segmentation and scoring of COVID-19 lung ultrasound images," *The Journal of The Acoustical Society of America*, vol. 150, no. 6, pp. 4118–4127, 2021. [Online]. Available: https://doi.org/10.1121/10.0007272

[49] L. Demi, M. Demi, A. Smargiassi, R. Inchingolo, F. Faita, and G. Soldati, "Ultrasonography in lung pathologies: new perspectives," *Multidisciplinary respiratory medicine*, vol. 9, no. 1, p. 27, May 2014. [Online]. Available: https://doi.org/10.1186/2049-6958-9-27

[50] G. Soldati, M. Demi, R. Inchingolo, A. Smargiassi, and L. Demi, "On the Physical Basis of Pulmonary Sonographic Interstitial Syndrome," *Journal of Ultrasound in Medicine*, vol. 35, no. 10, pp. 2075–2086, October 2016. [Online]. Available: https://doi.org/10.7863/ultra.15.08023

[51] G. Soldati, A. Smargiassi, A. A. Mariani, and R. Inchingolo, "Novel aspects in diagnostic approach to respiratory patients: Is it the time for a new semiotics?" *Multidisciplinary Respiratory Medicine*, vol. 12, no. 1, 2017. [Online]. Available: https://doi.org/10.1186/s40248-017-0098-z

[52] A. Smargiassi, R. Inchingolo, M. Chiappetta, L. P. Ciavarella, S. Lopatriello, G. M. Corbo, S. Margaritora, and L. Richeldi, "Agreement between chest ultrasonography and chest X-ray in patients who have undergone thoracic surgery: Preliminary results," *Multidisciplinary Respiratory Medicine*, vol. 14, no. 1, 2019. [Online]. Available: https://doi.org/10.1186/s40248-019-0171-x

[53] A. Smargiassi, R. Inchingolo, L. Calandriello, F. Lombardi, A. Calabrese, M. Siciliano, A. Larici, L. Demi, L. Richeldi, and G. Soldati, "Possible Role of Chest Ultrasonography for the Evaluation of Peripheral Fibrotic Pulmonary Changes in Patients Affected by Idiopathic Pulmonary Fibrosis-Pilot Case Series," *Applied Sciences*, vol. 10, March 2020. [Online]. Available: https://doi.org/10.3390/app10051617

[54] G. Soldati, A. Smargiassi, L. Demi, and R. Inchingolo, "Artifactual Lung Ultrasonography: It Is a Matter of Traps, Order, and Disorder," p. 1570, 2020. [Online]. Available: https://doi.org/10.3390/app10051570

[55] L. Demi, W. van Hoeve, R. J. G. van Sloun, G. Soldati, and M. Demi, "Determination of a potential quantitative measure of the state of the lung using lung ultrasound spectroscopy," *Scientific Reports*, vol. 7, no. 1, p. 12746, 2017. [Online]. Available: https://doi.org/10.1038/s41598-017-13078-9

[56] K. Mohanty, J. Blackwell, T. Egan, and M. Muller, "Characterization of the Lung Parenchyma Using Ultrasound Multiple Scattering," *Ultrasound in Medicine and Biology*, vol. 43, no. 5, pp. 993–1003, May 2017. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2017.01.011

[57] X. Zhang, T. Osborn, B. Zhou, D. Meixner, R. R. Kinnick, B. Bartholmai, J. F. Greenleaf, and S. Kalra, "Lung Ultrasound Surface Wave Elastography: A Pilot Clinical Study," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 9, pp. 1298–1304, 2017. [Online]. Available: https://doi.org/10.1109/tuffc.2017.2707981

[58] L. Demi, M. Demi, R. Prediletto, and G. Soldati, "Real-time multi-frequency ultrasound imaging for quantitative lung ultrasound – first clinical results," *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. 998–1006, 2020. [Online]. Available: https://doi.org/10.1121/10.0001723

[59] F. Mento, G. Soldati, R. Prediletto, M. Demi, and L. Demi, "Quantitative Lung Ultrasound Spectroscopy Applied to the Diagnosis of Pulmonary Fibrosis: The First Clinical Study," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2265–2273, 2020. [Online]. Available: https://doi.org/10.1109/tuffc.2020.3012289

[60] F. Mento and L. Demi, "On the Influence of Imaging Parameters on Lung Ultrasound B-line Artifacts, in vitro study," *Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. 975–983, 2020. [Online]. Available: https://doi.org/10.1121/10.0001797

[61] B. M. Wiley, B. Zhou, G. Pandompatam, J. Zhou, H. O. Kucuk, and X. Zhang, "Lung Ultrasound Surface Wave Elastography for Assessing Patients with Pulmonary Edema," *IEEE Transactions on Biomedical Engineering*, 2021. [Online]. Available: https://doi.org/10.1109/tbme.2021.3072891

[62] S. L. Contreras-Ojeda, J. A. Domínguez-Jiménez, and S. H. Contreras-Ortiz, "Analysis and Classification of Lung and Muscular Tissues in Ultrasound Images using 2D Wavelet Transform and Machine Learning," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 11583, 2020. [Online]. Available: https://doi.org/10.1117/12.2576368

[63] N. Anantrasirichai, W. Hayes, M. Allinovi, D. Bull, and A. Achim, "Line Detection as an Inverse Problem: Application to Lung Ultrasound Imaging," *IEEE Transactions on Medical Imaging*, vol. 36, no. 10, pp. 2045–2056, 2017. [Online]. Available: http://dx.doi.org/10.1109/TMI.2017.2715880

[64] R. Moshavegh, K. L. Hansen, H. Moller-Sorensen, M. B. Nielsen, and J. A. Jensen, "Automatic Detection of B-Lines in In Vivo Lung Ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 2, pp. 309–317, 2019. [Online]. Available: https://doi.org/10.1109/tuffc.2018.2885955

[65] O. Karakus, N. Anantrasirichai, A. Aguersif, S. Silva, A. Basarab, and A. Achim, "Detection of Line Artifacts in Lung Ultrasound Images of COVID-19 Patients Via Nonconvex Regularization," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2218–2229, 2020. [Online]. Available: https://doi.org/10.1109/tuffc.2020.3016092

[66] T. Kameda, N. Kamiyama, H. Kobayashi, Y. Kanayama, and N. Taniguchi, "Ultrasonic B-Line–Like Artifacts Generated with Simple Experimental Models Provide Clues to Solve Key Issues in B-Lines," *Ultrasound in Medicine and Biology*, vol. 45, no. 7, pp. 1617–1626, 2019. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2019.03.003

[67] F. Mento and L. Demi, "Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: An in vitro study," *The Journal of the Acoustical Society of America*, vol. 150, no. 6, pp. 4075–4082, 2021. [Online]. Available: https://doi.org/10.1121/10.0007482

[68] C. Brusasco, G. Santori, G. Tavazzi, G. Via, C. Robba, L. Gargani, F. Mojoli, S. Mongodi, E. Bruzzo, R. Trò, P. Boccacci, A. Isirdi, F. Forfori, F. Corradi, R. Biagini, A. Calvaruso, D. Costanzo, F. Cundari, A. Della Rocca, A. Laffi, E. Marchi, F. Marino, M. Monfroni, N. Mongiovì, S. Narteni, C. Piagnani, S. Prizio, L. Protsyak, M. Romani, I. Soldati, E. Taddei, L. Tecchi, and D. Trunfio, "Second-order grey-scale texture analysis of pleural ultrasound images to differentiate acute respiratory distress syndrome and cardiogenic pulmonary edema," *Journal of Clinical Monitoring and Computing*, 2020. [Online]. Available: https://doi.org/10.1007/s10877-020-00629-1

[69] S. Kulhare, X. Zheng, C. Mehanian, C. Gregory, M. Zhu, K. Gregory, H. Xie, J. McAndrew Jones, and B. Wilson, "Ultrasound-based detection of lung abnormalities using single shot detection convolutional neural networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11042 LNCS, pp. 65–73, 2018. [Online]. Available: https://doi.org/10.1007/978-3-030-01045-4_8

[70] R. J. G. Van Sloun and L. Demi, "Localizing B-Lines in Lung Ultrasonography by Weakly Supervised Deep Learning, In-Vivo Results," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 957–964, 2020. [Online]. Available: https://doi.org/10.1109/jbhi.2019.2936151

[71] H. Kerdegari, P. T. H. Nhat, A. McBride, R. Razavi, N. V. Hao, L. Thwaites, S. Yacoub, and A. Gomez, "Automatic detection of B-lines in lung ultrasound videos from severe dengue patients," in *Proceedings - International Symposium on Biomedical Imaging*, vol. 2021-April, 2021, pp. 989–993. [Online]. Available: https://doi.org/10.1109/ISBI48211.2021.9434006

[72] C. Baloescu, G. Toporek, S. Kim, K. McNamara, R. Liu, M. M. Shaw, R. L. McNamara, B. I. Raju, and C. L. Moore, "Automated Lung Ultrasound B-Line Assessment Using a Deep Learning Algorithm," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2312–2320, 2020. [Online]. Available: https://doi.org/10.1109/tuffc.2020.3002249

[73] X. Wang, J. S. Burzynski, J. Hamilton, P. S. Rao, W. F. Weitzel, and J. L. Bull, "Quantifying lung ultrasound comets with a convolutional neural network: Initial clinical results," *Computers in Biology and Medicine*, vol. 107, pp. 39–46, 2019. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2019.02.002

[74] C. Mehanian, S. Kulhare, R. Millin, X. Zheng, C. Gregory, M. Zhu, H. Xie, J. Jones, J. Lazar, A. Halse, T. Graham, M. Stone, K. Gregory, and B. Wilson, "Deep Learning-Based Pneumothorax Detection in Ultrasound Videos," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11798 LNCS, pp. 74–82, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-32875-7_9

[75] M. Jaščur, M. Bundzel, M. Malík, A. Dzian, N. Ferenčík, and F. Babič, "Detecting the absence of lung sliding in lung ultrasounds using deep learning," *Applied Sciences (Switzerland)*, vol. 11, no. 15, 2021. [Online]. Available: https://doi.org/10.3390/app11156976

[76] C.-H. Tsai, J. van der Burgt, D. Vukovic, N. Kaur, L. Demi, D. Canty, A. Wang, A. Royse, C. Royse, K. Haji, J. Dowling, G. Chetty, and D. Fontanarosa, "Automatic deep learning-based pleural effusion classification in lung ultrasound images for respiratory pathology diagnosis," *Physica Medica*, vol. 83, pp. 38–45, 2021. [Online]. Available: https://doi.org/10.1016/j.ejmp.2021.02.023

[77] B. Zhou and X. Zhang, "Lung mass density analysis using deep neural network and lung ultrasound surface wave elastography," *Ultrasonics*, vol. 89, pp. 173–177, 2018. [Online]. Available: https://doi.org/10.1016/j.ultras.2018.05.011

[78] B. Zhou, B. J. Bartholmai, S. Kalra, and X. Zhang, "Predicting lung mass density of patients with interstitial lung disease and healthy subjects using deep neural network and lung ultrasound surface wave elastography," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 104, 2020. [Online]. Available: https://doi.org/10.1016/j.jmbbm.2020.103682

[79] B. Zhou, B. J. Bartholmai, S. Kalra, T. Osborn, and X. Zhang, "Lung mass density prediction using machine learning based on ultrasound surface wave elastography and pulmonary function testing," *Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 1318–1323, 2021. [Online]. Available: https://doi.org/10.1121/10.0003575

[80] A. Almeida, A. Bilbao, L. Ruby, M. B. Rominger, D. Lopez-De-Ipina, J. Dahl, A. Elkaffas, and S. J. Sanabria, "Lung ultrasound for point-of-care COVID-19 pneumonia stratification: Computer-aided diagnostics in a smartphone. First experiences classifying semiology from public datasets," in *IEEE International Ultrasonics Symposium, IUS*, vol. 2020-Septe, 2020. [Online]. Available: https://doi.org/10.1109/IUS46767.2020.9251716

[81] D. Hou, R. Hou, and J. Hou, "Interpretable Saab Subspace Network for COVID-19 Lung Ultrasound Screening," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2020*, 2020, pp. 393–398. [Online]. Available: https://doi.org/10.1109/UEMCON51285.2020.9298069

[82] S. Erfanian Ebadi, D. Krishnaswamy, S. E. S. Bolouri, D. Zonoobi, R. Greiner, N. Meuser-Herr, J. L. Jaremko, J. Kapur, M. Noga, and K. Punithakumar, "Automated detection of pneumonia in lung ultrasound using deep video classification for COVID-19," *Informatics in Medicine Unlocked*, vol. 25, 2021. [Online]. Available: https://doi.org/10.1016/j.imu.2021.100687

[83] U. Khan, F. Mento, L. N. Giacomaz, R. Trevisan, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "Deep learning-based classification of reduced lung ultrasound data from covid-19 patients," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 5, pp. 1661–1669, March 2022. [Online]. Available: https://doi.org/10.1109/tuffc.2022.3161716

[84] M. La Salvia, G. Secco, E. Torti, G. Florimbi, L. Guido, P. Lago, F. Salinaro, S. Perlini, and F. Leporati, "Deep learning and lung ultrasound for Covid-19 pneumonia detection and severity classification," *Computers in Biology and Medicine*, vol. 136, 2021. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2021.104742

[85] G. Soldati, V. Giunta, S. Sher, F. Melosi, and C. Dini, ""Synthetic" Comets: A New Look at Lung Sonography," *Ultrasound in Medicine and Biology*, vol. 37, no. 11, pp. 1762–1770, 2011. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2011.05.024

[86] G. Soldati, R. Inchingolo, A. Smargiassi, S. Sher, R. Nenna, C. D. Inchingolo, and S. Valente, "Ex Vivo Lung Sonography: Morphologic-Ultrasound Relationship," *Ultrasound in Medicine and Biology*, vol. 38, no. 7, pp. 1169–1179, 2012. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2012.03.001

[87] G. Soldati, A. Smargiassi, R. Inchingolo, S. Sher, R. Nenna, S. Valente, C. D. Inchingolo, and G. M. Corbo, "Lung ultrasonography may provide an indirect estimation of lung porosity and airspace geometry," *Respiration*, vol. 88, no. 6, pp. 458–468, 2014. [Online]. Available: https://doi.org/10.1159/000368086

[88] L. Demi, T. Egan, and M. Muller, "Lung ultrasound imaging, a technical review," *Applied Sciences (Switzerland)*, vol. 10, no. 2, 2020. [Online]. Available: https://doi.org/10.3390/app10020462

[89] K. Mohanty, Y. Karbalaeisadegh, J. Blackwell, M. Ali, B. Masuodi, T. Egan, and M. Muller, "In Vivo Assessment of Pulmonary Fibrosis and Pulmonary Edema in Rodents Using Ultrasound Multiple Scattering," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2274–2280, 2020. [Online]. Available: https://doi.org/10.1109/TUFFC.2020.3023611

[90] R. Roshankhah, J. Blackwell, M. H. Ali, B. Masuodi, T. Egan, and M. Muller, "Detecting pulmonary nodules by using ultrasound multiple scattering," *The Journal of the Acoustical Society of America*, vol. 150, no. 6, pp. 4095–4102, 2021. [Online]. Available: https://doi.org/10.1121/10.0006666

[91] T. H. Lye, R. Roshankhah, Y. Karbalaeisadegh, S. A. Montgomery, T. M. Egan, M. Muller, and J. Mamou, "In vivo assessment of pulmonary fibrosis and edema in rodents using the backscatter coefficient and envelope statistics," *The Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 183–192, July 2021. [Online]. Available: https://doi.org/10.1121/10.0005481

[92] J. Zhou and X. Zhang, "A Lung Phantom Model to Study Pulmonary Edema Using Lung Ultrasound Surface Wave Elastography," *Ultrasound in Medicine and Biology*, vol. 44, no. 11, pp. 2400–2405, 2018. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2018.06.008

[93] M. Demi, "The impact of multiple concurrent factors on the length of the ultrasound pulmonary vertical artifacts as illustrated through the experimental and numerical analysis of simple models," *The Journal of the Acoustical Society of America*, vol. 150, no. 3, pp. 2106–2115, 2021. [Online]. Available: https://doi.org/10.1121/10.0006413

[94] T. Kameda, N. Kamiyama, and N. Taniguchi, "The Mechanisms Underlying Vertical Artifacts in Lung Ultrasound and Their Proper Utilization for the Evaluation of Cardiogenic Pulmonary Edema," 2022. [Online]. Available: https://doi.org/10.3390/diagnostics12020252

[95] F. Wolfram, C. Braun, H. Gutsche, and T. G. Lesser, "In Vivo Assessment of Lung Ultrasound Features Mimicking Viral Pneumonia Using a Large Animal Model," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2258–2264, 2020. [Online]. Available: https://doi.org/10.1109/TUFFC.2020.3010299

[96] E. Peschiera, F. Mento, and L. Demi, "Numerical study on lung ultrasound B-line formation as a function of imaging frequency and alveolar geometries," *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2304–2311, April 2021. [Online]. Available: https://doi.org/10.1121/10.0003930

[97] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, no. 2, pp. 1–12, March 2010. [Online]. Available: https://doi.org/10.1117/1.3360308

[98] O. Ostras, D. E. Soulioti, and G. Pinton, "Diagnostic ultrasound imaging of the lung: A simulation approach based on propagation and reverberation in the human body," *The Journal of the Acoustical Society of America*, vol. 150, no. 5, pp. 3904–3913, 2021. [Online]. Available: https://doi.org/10.1121/10.0007273

[99] X. Zhang, B. Zhou, T. Osborn, B. Bartholmai, and S. Kalra, "Lung Ultrasound Surface Wave Elastography for Assessing Interstitial Lung Disease," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1346–1352, 2019. [Online]. Available: https://dx.doi.org/10.1109%2FTBME.2018.2872907

[100] X. Zhang, B. Zhou, B. Bartholmai, S. Kalra, and T. Osborn, "A quantitative method for measuring the changes of lung surface wave speed for assessing disease progression of interstitial lung disease," *Ultrasound in Medicine and Biology*, vol. 45, no. 3, pp. 741–748, 2019. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2018.11.003

[101] M. Demi, R. Prediletto, G. Soldati, and L. Demi, "Physical Mechanisms Providing Clinical Information from Ultrasound Lung Images: Hypotheses and Early Confirmations," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 3, pp. 612–623, 2020. [Online]. Available: https://doi.org/10.1109/tuffc.2019.2949597

[102] World Health Organization (WHO), "WHO Coronavirus Disease (COVID-19) Dashboard." [Online]. Available: https://covid19.who.int/

[103] World Health Organization, "What are the symptoms of COVID-19?" 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses

[104] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, and G. Gao, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020. [Online]. Available: https://doi.org/10.1056/nejmoa2001017

[105] World Health Organization, "Laboratory testing for coronavirus disease (COVID-19) in suspected human cases," 2020. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/331501/WHO-COVID-19-laboratory-2020.5-eng.pdf?sequence=1&isAllowed=y

[106] S. Woloshin, N. Patel, and A. S. Kesselheim, "False Negative Tests for SARS-CoV-2 Infection - Challenges and Implications," *The New England journal of medicine*, vol. 383, no. 6, p. e38, 2020. [Online]. Available: https://doi.org/10.1056/NEJMp2015897

[107] B. Xu, Y. Xing, J. Peng, Z. Zheng, W. Tang, Y. Sun, C. Xu, and F. Peng, "Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy," *European Radiology*, vol. 30, no. 10, pp. 5720–5727, 2020. [Online]. Available: https://doi.org/10.1007/s00330-020-06934-2

[108] E. Picano and P. A. Pellikka, "Ultrasound of extravascular lung water: a new standard for pulmonary congestion," *European Heart Journal*, vol. 37, no. 27, pp. 2097–2104, May 2016. [Online]. Available: https://doi.org/10.1093/eurheartj/ehw164

[109] ICLUS Team and BlueTensor, "ICLUS web application." [Online]. Available: https://iclusweb.bluetensor.ai/

[110] E. Poggiali, A. Dacrema, D. Bastoni, V. Tinelli, E. Demichele, P. Mateo Ramos, T. Marcianò, M. Silva, A. Vercelli, and A. Magnacavallo, "Can Lung US Help Critical Care Clinicians in the Early Diagnosis of Novel Coronavirus (COVID-19) Pneumonia?" *Radiology*, vol. 295, no. 3, pp. E6–E6, March 2020. [Online]. Available: https://doi.org/10.1148/radiol.2020200847

[111] P. Lomoro, F. Verde, F. Zerboni, I. Simonetti, C. Borghi, C. Fachinetti, A. Natalizi, and A. Martegani, "COVID-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and CT: single-center study and comprehensive radiologic literature review," *European Journal of Radiology Open*, vol. 7, January 2020. [Online]. Available: https://doi.org/10.1016/j.ejro.2020.100231

[112] A. Nouvenne, A. Ticinesi, A. Parise, B. Prati, M. Esposito, V. Cocchi, E. Crisafulli, A. Volpi, S. Rossi, E. G. Bignami, M. Baciarello, E. Brianti, M. Fabi, and T. Meschi, "Point-of-Care Chest Ultrasonography as a Diagnostic Resource for COVID-19 Outbreak in Nursing Homes," *Journal of the*

American Medical Directors Association, vol. 21, no. 7, pp. 919–923, July 2020. [Online]. Available: https://doi.org/10.1016/j.jamda.2020.05.050

[113] K. Yasukawa and T. Minami, "Point-of-Care Lung Ultrasound Findings in Patients with COVID-19 Pneumonia," *The American Journal of Tropical Medicine and Hygiene*, vol. 102, no. 6, pp. 1198–1202, 2020. [Online]. Available: https://doi.org/10.4269/ajtmh.20-0280

[114] C. Xing, Q. Li, H. Du, W. Kang, J. Lian, and L. Yuan, "Lung ultrasound findings in patients with COVID-19 pneumonia," *Critical Care*, vol. 24, no. 1, p. 174, 2020. [Online]. Available: https://doi.org/10.1186/s13054-020-02876-9

[115] Q.-Y. Peng, X.-T. Wang, L.-N. Zhang, and C. C. C. U. S. G. (CCUSG), "Findings of lung ultrasonography of novel corona virus pneumonia during the 2019–2020 epidemic," *Intensive Care Medicine*, vol. 46, no. 5, pp. 849–850, 2020. [Online]. Available: https://doi.org/10.1007/s00134-020-05996-6

[116] G. Duclos, A. Lopez, M. Leone, and L. Zieleskiewicz, ""No dose" lung ultrasound correlation with "low dose" CT scan for early diagnosis of SARS-CoV-2 pneumonia," *Intensive Care Medicine*, vol. 46, no. 6, pp. 1103–1104, 2020. [Online]. Available: https://doi.org/10.1007/s00134-020-06058-7

[117] M. Yuan, W. Yin, Z. Tao, W. Tan, and Y. Hu, "Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China," *PLOS ONE*, vol. 15, no. 3, p. e0230548, March 2020. [Online]. Available: https://doi.org/10.1371/journal.pone.0230548

[118] F. Mento, T. Perrone, A. Fiengo, F. Tursi, V. N. Macioce, A. Smargiassi, R. Inchingolo, and L. Demi, "Limiting the areas inspected by lung ultrasound leads to an underestimation of COVID-19 patients' condition," *Intensive Care Medicine*, vol. 47, no. 7, pp. 811–812, 2021. [Online]. Available: https://doi.org/10.1007/s00134-021-06407-0

[119] S. H. Yoon, K. H. Lee, J. Y. Kim, Y. K. Lee, H. Ko, K. H. Kim, C. M. Park, and Y. H. Kim, "Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea," *Korean journal of radiology*, vol. 21, no. 4, pp. 494–500, April 2020. [Online]. Available: https://doi.org/10.3348/kjr.2020.0132

[120] Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song, "Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review," *European Radiology*, pp. 1–9, March 2020. [Online]. Available: https://dx.doi.org/10.1007%2Fs00330-020-06801-0

[121] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li, S. Li, H. Shan, A. Jacobi, and M. Chung, "Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection," *Radiology*, vol. 295, no. 3, p. 200463, February 2020. [Online]. Available: https://doi.org/10.1148/radiol.2020200463

[122] L. Gargani, M. Doveri, L. D'Errico, F. Frassi, M. L. Bazzichi, A. D. Sedie, M. C. Scali, S. Monti, S. Mondillo, S. Bombardieri, D. Caramella, and E. Picano, "Ultrasound lung comets in systemic sclerosis: a chest sonography hallmark of pulmonary interstitial fibrosis," *Rheumatology*, vol. 48, no. 11, pp. 1382–1387, August 2009. [Online]. Available: https://doi.org/10.1093/rheumatology/kep263

[123] F. Corradi, C. Brusasco, A. Vezzani, G. Santori, T. Manca, L. Ball, F. Nicolini, T. Gherli, and V. Brusasco, "Computer-Aided Quantitative Ultrasonography for Detection of Pulmonary Edema in Mechanically Ventilated Cardiac Surgery Patients," *CHEST*, vol. 150, no. 3, pp. 640–651, September 2016. [Online]. Available: https://doi.org/10.1016/j.chest.2016.04.013

[124] L. Gargani, "Lung ultrasound: a new tool for the cardiologist," *Cardiovascular ultrasound*, vol. 9, p. 6, February 2011. [Online]. Available: https://doi.org/10.1186/1476-7120-9-6

[125] ——, "Ultrasound of the Lungs: More than a Room with a View," *Heart Failure Clinics*, vol. 15, no. 2, pp. 297–303, 2019. [Online]. Available: https://doi.org/10.1016/j.hfc.2018.12.010

[126] P. Tortoli, L. Bassi, E. Boni, A. Dallai, F. Guidi, and S. Ricci, "ULA-OP: An Advanced Open Platform for Ultrasound Research," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, pp. 2207–2216, October 2009. [Online]. Available: https://doi.org/10.1109/tuffc.2009.1303

[127] J. R. Cook, R. R. Bouchard, and S. Y. Emelianov, "Tissue-mimicking phantoms for photoacoustic and ultrasonic imaging," *Biomedical optics express*, vol. 2, no. 11, pp. 3193–3206, November 2011. [Online]. Available: https://dx.doi.org/10.1364%2FBOE.2.003193

[128] L. Spicci, "FEM Simulation for 'Pulse-Echo' Performances of an Ultrasound Imaging Linear Probe," *Proceedings of the COMSOL 2013 Rotterdam*, 2013. [Online]. Available: https://www.comsol.it/paper/download/181457/spicci_paper.pdf

[129] T. L. Szabo, "Chapter 4 - Attenuation," in *Diagnostic Ultrasound Imaging: Inside Out (Second Edition)*, T. Szabo, Ed. Boston: Academic Press, 2014, pp. 81–119. [Online]. Available: https://doi.org/10.1016/B978-0-12-396487-8.00004-5

[130] World Health Organization (WHO), "World Health Statistics 2016." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[131] ——, "WHO Coronavirus Disease (COVID-19) Dashboard." [Online]. Available: https://covid19.who.int/

[132] M. Demi, G. Soldati, and L. Demi, "On the artefactual information of ultrasound lung images: A lines and B lines," *Proceedings of Meetings on Acoustics*, vol. 35, no. 1, p. 20003, November 2018. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/2.0000943

[133] D. A. Lichtenstein, G. A. Mezière, J. F. Lagoueyte, P. Biderman, I. Goldstein, and A. Gepner, "A-lines and B-lines: Lung ultrasound as a bedside tool for predicting pulmonary artery occlusion pressure in the critically ill," *Chest*, 2009. [Online]. Available: https://doi.org/10.1378/chest.09-0001

[134] E. Peschiera, F. Mento, and L. Demi, "The data resulting from the described simulations." [Online]. Available: https://sites.google.com/view/drlibertariodemi/data?authuser=0

[135] Forum of International Respiratory Societies, "The Global Impact of Respiratory Disease- Second Edition," European Respiratory Society, Sheffield, Tech. Rep., 2017.

[136] World Health Organization, "Coronavirus disease 2019 Situation Report – 128," 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200527-covid-19-sitrep-128.pdf?sfvrsn=11720c0a_2

[137] J. H. Ryu, T. Moua, C. E. Daniels, T. E. Hartman, E. S. Yi, J. P. Utz, and A. H. Limper, "Idiopathic Pulmonary Fibrosis: Evolving Concepts," *Mayo Clinic Proceedings*, vol. 89, no. 8, pp. 1130–1142, 2014. [Online]. Available: https://doi.org/10.1016/j.mayocp.2014.03.016

[138] F. J. Martinez, H. R. Collard, A. Pardo, G. Raghu, L. Richeldi, M. Selman, J. J. Swigris, H. Taniguchi, and A. U. Wells, "Idiopathic pulmonary fibrosis," *Nature Reviews Disease Primers*, vol. 3, no. 1, p. 17074, 2017. [Online]. Available: https://doi.org/10.1038/nrdp.2017.74

[139] J. M. Oldham and I. Noth, "Idiopathic pulmonary fibrosis: early detection and referral," *Respiratory medicine*, vol. 108, no. 6, pp. 819–829, June 2014. [Online]. Available: https://doi.org/10.1016/j.rmed.2014.03.008

[140] G. Soldati, "2.17 - Biomedical Applications of Ultrasound," in *Comprehensive Biomedical Physics*, A. B. T. C. B. P. Brahme, Ed. Oxford: Elsevier, 2014, pp. 401–436. [Online]. Available: https://doi.org/10.1016/B978-0-444-53632-7.00219-7

[141] D. A. Lichtenstein and G. A. Mezière, "Relevance of Lung Ultrasound in the Diagnosis of Acute Respiratory Failure*: The BLUE Protocol," *Chest*, vol. 134, no. 1, pp. 117–125, 2008. [Online]. Available: https://doi.org/10.1378/chest.07-2800

[142] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press, 2000. [Online]. Available: https://doi.org/10.1017/CBO9780511801389

[143] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Boca Raton, FL: CRC Press, 1984.

[144] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, "COVID-19 Detection through Transfer Learning Using Multimodal Imaging Data," *IEEE Access*, vol. 8, pp. 149 808–149 824, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.3016780

[145] J. Diaz-Escobar, N. E. Ordóñez-Guillén, S. Villarreal-Reyes, A. Galaviz-Mosqueda, V. Kober, R. Rivera-Rodriguez, and J. E. Lozano Rizk, "Deep-learning based detection of COVID-19 using lung ultrasound imagery," *PLoS ONE*, vol. 16, no. 8 August, 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0255886

[146] N. Awasthi, A. Dayal, L. R. Cenkeramaddi, and P. K. Yalavarthy, "Mini-COVIDNet: Efficient Lightweight Deep Neural Network for Ultrasound Based Point-of-Care Detection of COVID-19," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 6, pp. 2023–2037, 2021. [Online]. Available: https://doi.org/10.1109/TUFFC.2021.3068190

[147] F. Mento, U. Khan, F. Faita, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "State of The Art in Lung Ultrasound, Shifting From Qualitative To Quantitative Analyses," *Ultrasound in Medicine and Biology*, in press, 2022. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2022.07.007

[148] 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, and L. Demi, "On Lung Ultrasound Patterns Specificity in the Management of COVID-19 Patients," *Journal of Ultrasound in Medicine*, vol. 39, no. 11, pp. 2283–2284, May 2020. [Online]. Available: https://doi.org/10.1002/jum.15326

[149] F. Mento, G. Soldati, R. Prediletto, M. Demi, and L. Demi, "Differentiation of Pulmonary Fibrosis by Means of Quantitative Lung Ultrasound Spectroscopy, First Clinical Study in Humans," in *2020 IEEE International Ultrasonics Symposium (IUS)*, 2020, pp. 1–4. [Online]. Available: https://doi.org/10.1109/IUS46767.2020.9251740

[150] F. Mento and L. Demi, "Effect of Imaging Parameters on the Visualization of Lung Ultrasound B-line Artifacts," in *2020 IEEE International Ultrasonics Symposium (IUS)*, 2020, pp. 1–4. [Online]. Available: https://doi.org/10.1109/IUS46767.2020.9251818

[151] D. Yaron, D. Keidar, E. Goldstein, Y. Shachar, A. Blass, O. Frank, N. Schipper, N. Shabshin, A. Grubstein, D. Suhami, N. R. Bogot, C. S. Weiss, E. Sela, A. A. Dror, M. Vaturi, F. Mento, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, T. Perrone, L. Demi, M. Galun, S. Bagon, Y. M. Elyada, and Y. C. Eldar, "Point of Care Image Analysis for COVID-19," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8153–8157. [Online]. Available: https://doi.org/10.1109/ICASSP39728.2021.9413687

[152] F. Mento, T. Perrone, A. Fiengo, V. N. Macioce, F. Tursi, A. Smargiassi, R. Inchingolo, and L. Demi, "A Multicenter Study Assessing Artificial Intelligence Capability in Scoring Lung Ultrasound Videos of COVID-19 Patients," in *2021 IEEE International Ultrasonics Symposium (IUS)*, 2021, pp. 1–3. [Online]. Available: https://doi.org/10.1109/IUS52206.2021.9593821

[153] F. Mento and L. Demi, "Impact of Frequency, Bandwidth, Focus, and Angle of Incidence on Lung Ultrasound Vertical Artifacts' Intensity, in-vitro Study," in *2021 IEEE International Ultrasonics Symposium (IUS)*, 2021, pp. 1–3. [Online]. Available: https://doi.org/10.1109/IUS52206.2021.9593821

[154] U. Khan, F. Mento, L. N. Giacomaz, R. Trevisan, A. Smargiassi, R. Inchingolo, T. Perrone, and L. Demi, "Impact of pixel, intensity, and temporal resolution on automatic scoring of LUS from Coronavirus disease 2019 patients," *Proceedings of Meetings on Acoustics*, in press, 2022.

[155] L. L. Custode, F. Mento, S. Afrakhteh, F. Tursi, A. Smargiassi, R. Inchingolo, T. Perrone, G. Iacca, and L. Demi, "Neuro-symbolic interpretable AI for automatic COVID-19 patient-stratification based on standardised lung ultrasound data," *Proceedings of Meetings on Acoustics*, in press, 2022.