Computers in Biology and Medicine 200 (2026) 111365

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Fully automated Quantitative Lung Ultrasound spectroscopy for the differential diagnosis of lung diseases: The first multicenter in-vivo clinical study

Mattia Perpenti [a], Federico Mento [a], Giovanni Pierro [b], Alessandro Perrotta [b], Tiziano Perrone [c], Andrea Smargiassi [b], Riccardo Inchingolo [b], Libertario Demi [a,*]

[a] Department of Information Engineering and Computer Science, University of Trento, Trento, 38123, Italy
[b] Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, 00168, Italy
[c] Department of Internal Medicine, Humanitas Gavazzeni, Bergamo, 24125, Italy

ARTICLE INFO

Keywords:
Automatic segmentation algorithm
Cardiogenic pulmonary edema (CPE)
Chronic obstructive pulmonary disease (COPD)
Lung ultrasound (LUS) spectroscopy
Pleural line (PL)
Pneumonia (PNE)
Pulmonary fibrosis (PF)
Pulmonary fibrosis exacerbated (PFE)

ABSTRACT

Quantitative Lung Ultrasound (LUS) analyses are demonstrating significant improvements in the diagnoses of lung diseases. However, recent approaches are limited to the manual segmentation of artifactual patterns (i.e., vertical artifacts), whose detection is subjective, dependent on imaging settings, and time-consuming. These limitations can be partially addressed by segmenting anatomical patterns [e.g., the pleural line (PL)], which are less affected by variations in imaging settings. Additionally, an automated segmentation algorithm would reduce segmentation time and subjectivity. In this study, we proposed an Automatic Segmentation Algorithm of the Pleural line (ASAP). ASAP was first optimized on one dataset [composed of 52 patients affected by Chronic Obstructive Pulmonary Disease (COPD), Pulmonary Fibrosis (PF), and Pulmonary Fibrosis Exacerbated (PFE)] and then tested on an external dataset [composed of 34 patients affected by Pneumonia (PNE) and Cardiogenic Pulmonary Edema (CPE)]. In total, 12761 multifrequency radiofrequency images from 86 patients were segmented into three regions: pre-pleura, pleura, and post-pleura. These regions were quantified with five I_{TOT_R} parameters (where R = 1, 2, 3, 4, and 5), each characterized with three features: Center Frequency, Bandwidth, and *max* I_{TOT_R}. The discrimination potential of these features was tested through different binary classifiers. Results show how ASAP segments in quasi real-time with 10 images/s and achieves a segmentation accuracy of 97.70%. The differential diagnosis of lung diseases shows accuracies up to 87.48, outperforming the state of the art.

1. Introduction

New horizons in the clinical diagnoses of pulmonary diseases led to Lung Ultrasound (LUS). LUS ability to provide real-time imaging with non-ionizing radiation makes it a valuable tool in medicine [1–3]. Nonetheless, LUS can only assess the lung surface and relies on the detection of imaging patterns [i.e., Pleural line (PL), vertical artifacts, horizontal artifacts, consolidations] [1–5]. The detection of these patterns is influenced by different imaging settings (i.e., frequency, dynamic range), subjective, and poorly reproducible [6,7]. As other lung diagnostic techniques implemented AI based solutions [8–12], to mitigate LUS flaws, LUS semi-quantitative AI-based approaches have been proposed in literature [13–25]. Although these approaches improved reproducibility, they are trained on ground-truths defined by subjective interpretations. Thus, they remain qualitative or semi-quantitative [26]. To further mitigate subjectivity and improve specificity, quantitative approaches have been developed [26,27]. Previous quantitative studies focused on the spectral characterization of vertical artifacts, whose presence is associated to diseases with a reduction of the air-spaces dimension [e.g., Pulmonary Fibrosis (PF), Pneumonia (PNE), and Cardiogenic Pulmonary Edema (CPE)] [28–34]. These studies demonstrated specificity improvements in differentiating these diseases [29–31]. However, vertical artifacts were manually segmented by human operators, affecting the method reproducibility [29–31]. Another study focused on the characterization of lung surface roughness by means of a multifrequency approach to indirectly estimates the size of alveolar structure along the lung surface [35]. This approach is particularly interesting for patients affected by diseases characterized by enlarged air spaces dimension [i.e., Chronic Obstructive Pulmonary

* Corresponding author.
E-mail address: libertario.demi@unitn.it (L. Demi).

https://doi.org/10.1016/j.compbiomed.2025.111365
Received 15 July 2025; Received in revised form 8 November 2025; Accepted 1 December 2025
Available online 10 December 2025

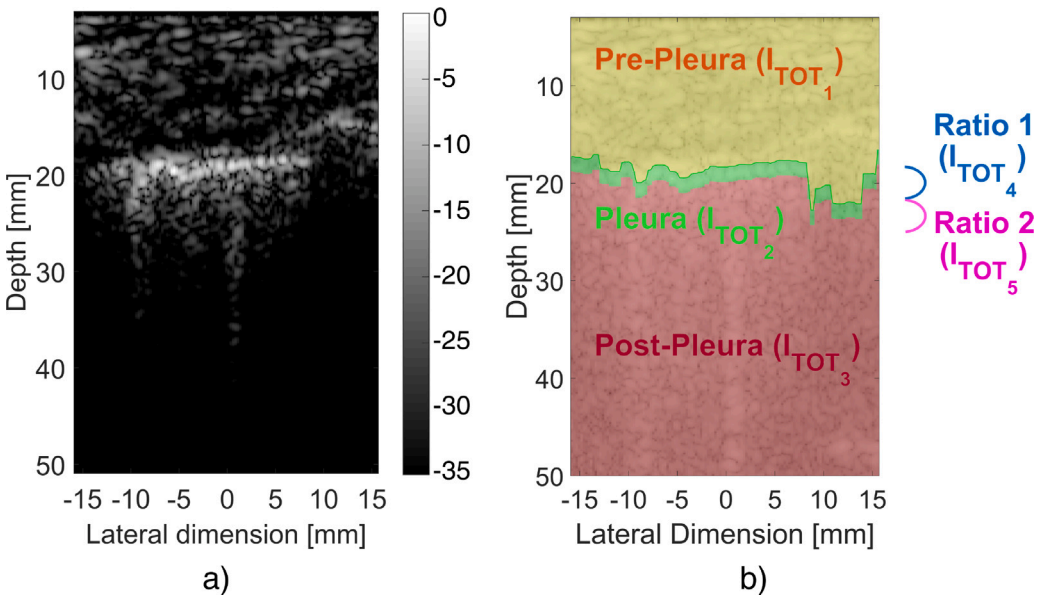

Fig. 1. (a) LUS Image displayed with a dynamic range of −35 dB. (b) Subdivision of the LUS image into the three regions: Pre-Pleura (I_{TOT_1}) (yellow), Pleura (I_{TOT_2}) (green), and Post-Pleura (I_{TOT_3}) (red); ratio 1 (I_{TOT_4}) (blue) and ratio 2 (I_{TOT_5}) (pink) are highlighted.

Diseases (COPD)] [35]. However, it was never tested *in-vivo*, where PL segmentation is necessary to analyze lung surface roughness.

In this study, we have three aims. Aim 1: the development of an Automatic Segmentation Algorithm of the Pleural line (ASAP) that operates in quasi real-time (AIM1). Aim 2: the validation of ASAP with respect to the state of the art (consisting in the manual segmentation) on one dataset (AIM2). Here ASAP was optimized with a ground-truth obtained from the manual segmentation of the PL. To simplify and accelerate the ground-truth generation, an assisted manual segmentation algorithm (AMSA) was developed. Aim 3: apply ASAP as the basis to perform a quantitative LUS spectroscopy for the differential diagnosis of lung diseases, test its performance on a second independent dataset and compare the final diagnostic results with the state of the art (AIM3) [31].

In this paper, we propose a quantitative approach that includes the spectral characterization of both artifactual and anatomical patterns of radiofrequency (RF) data. Three LUS image regions, pre-pleura, pleura, and post pleura (see Fig. 1), were characterized by center frequency, bandwidth, and total intensity to differentiate lung diseases. The efficiency of these features was successively tested through different binary classifiers [29,31,36]. This approach focuses on mitigating subjectivity and evaluates diseases both characterized by a reduction of air-spaces dimension (PF, CPE, and PNE) and by an enlargement of air-spaces (COPD). The final results demonstrate a fully automatic quantitative LUS approach that works in quasi real-time and improved classification performance for all the diseases investigated.

The article is organized as follows. The datasets, ASAP and the quantitative approach methodology are described in Section 2. The main results are presented in Section 3. To conclude, the observations and final conclusions are discussed in Section 4.

2. Materials and methods

2.1. Datasets

Two datasets composed of RF multifrequency data acquired in two italian clinical centers were utilized: the Lung Ultrasound Acute Respiratory Disease (LUSARD) dataset. acquired from Humanitas Gavazzeni Hospital of Bergamo and SAURON dataset, acquired from Fondazione Policlinico Universitario Agostino Gemelli of Rome. For both datasets, the acquisitions follow a multifrequency modality based on orthogonal sub-bands centered around specific center frequencies [29,37], and implemented with an ULtrasound Advanced Open Platform (ULA-OP), connected to linear probes [38]. As previous studies demonstrated, quantitative LUS spectroscopy applied to RF data from linear probes achieves higher performance than convex probes. Precisely, a LA533 linear probe (Esaote, Florence, Italy) [31] (192 elements, 245-μm pitch) was used for LUSARD and a LA523 (Esaote, Florence, Italy) (192 elements, 245-μm pitch) for SAURON. The orthogonal sub-bands selected were centered at 3, 4, 5, and 6 MHz. A 2-μs-time-length pulse was transmitted for each center frequency, guaranteeing, through different sets of number of cycles, a fixed band (1-MHz of bandwidth at −10 dB) [29,31,37]. The transfer functions of the probes were equalized to guarantee an equalized pressure output at PL [29]. The sub-aperture used to transmit and receive was set at 64-elements, generating images of the size of 129 lines (along the lateral dimension). Each line is independent and is generated by transmitting and receiving a 2-μs-time-length pulse (no spatial compounding). In total 129 lines were generated to form an image for each center frequency. The Pulse Repetition Frequency (PRF) was set at 4 KHz, corresponding to a frame rate of $\frac{PRF}{129} = 31$ Hz. The focal point was set at PL position (≈ 2 cm) with a maximum imaging depth of 6.5 cm. A dynamic sinc apodization function was applied. A sampling frequency of 50 MHz was utilized for digitization [31]. A speed of sound of 1546 m/s was assumed for the time-to-space conversion (along the depth) [29]. The time gain compensation (TGC) was set to approximately 1.77 dB/MHz cm [31].

2.1.1. LUSARD

This dataset was acquired within Emergency and Urgency Department, Humanitas Gavazzeni, Bergamo, Italy, in the context of LUSARD study (approved by the Independent Ethics Committee for Clinical Studies of IRCCS Clinical Institute Humanitas of Rozzano: protocol number 553/21). The dataset is composed of 34 patients, 17 diagnosed with PNE and 17 with CPE through standard clinical examinations [31]. Data were acquired considering 12 different areas. Only data displaying vertical artifacts were stored. A scan consists of a video of 23 multifrequency images, each composed by a set of 4 images of the same area generated with 4 different center frequencies. This dataset was used in a previous study, and would be used to benchmark the results of our analysis.

Table 1
Summary of the data selected from the two datasets as a number of patients and images.

LUSARD	Patients	Images
CPE	17	1426
PNE	17	1311
SAURON	Patients	Images
COPD	19	3550
PF	20	3944
PFE	12	2530

2.1.2. SAURON

This dataset is presented for the first time and represents an improved version of LUSARD. In fact, all the area were acquired independently from the presence of specific patterns in the image. As all the areas were scanned and data were saved, 10 areas were selected to facilitate clinical data acquisition. Is important to clarify that previous studied demonstrated how the reduction from 12 to 10 area does not compromise the quality of the analysis [39].

The dataset was acquired within Fondazione Policlinico Universitario Agostino Gemelli (approved by the Ethical Committee of the Fondazione Policlinico Universitario Gemelli IRCCS in Rome, protocol 0002865/22, ID 4710). A total of 51 patients grouped into three lung diseases were examined: 20 patients (38.5%) affected by PF (14 males and 6 females between 46 and 88 years with an average age of ≈73.6 years), 12 patients (23.0%) affected by Pulmonary Fibrosis Exacerbated (PFE) (8 males and 4 females between 54 and 87 years with an average age of ≈72.4 years), 20 patients (38.5%) affected by COPD (14 males and 6 females between 45 and 79 years with an average age of ≈68.2 years). Diagnoses were made through standard clinical examinations (e.g., CT).

2.1.3. Data selection

From LUSARD, interstitial pneumonia, lobar pneumonia, and COVID-19 pneumonia were aggregated as a unique disease: PNE. Successively, the pathologies most represented with linear probe data are selected, PNE and CPE (both having 17 patients). From SAURON, all the diseases were selected. The obtained dataset is summarized in Table 1. To start the analysis, the following preprocessing steps were applied to these data. Unwanted frequency components were removed by filtering the RF data with a Butterworth filter (1.8 MHz bandwidth) centered at the transmitted center frequency. Successively, the signal envelope was extracted through the Hilbert transform. Each image was then normalized with respect to its maximum value [29,31,36]. As done in previous studies, images were displayed in logarithmic scale with a 35-dB dynamic range [29,31,36].

2.2. Assisted manual segmentation algorithm (AMSA) for ground-truth definition

To properly quantify the pre-pleura, pleura, and post-pleura regions of LUS images, PL needs to be detected and segmented. This procedure is crucial and time-consuming, given that each RF image is usually manually segmented by an operator [36]. In this work, we perform an assisted manual segmentation to obtain three regions from each LUS image for the quantitative analysis, but also to generate a ground-truth for the segmented images. Consequently, all the 10,024 multifrequency images from SAURON dataset were manually segmented through AMSA as follows. From each multifrequency image, the image with the lowest frequency is selected (3 MHz), to consider the largest wavelength and thus the lowest spatial resolution. From this image, the segmentation starts by defining a series of $Npoints$ (N varies from image to image and is chosen by the operator) on PL (Fig. 2a). The initial and final points of PL are selected as the first and last visible points, respectively, from the left side of the lateral dimension. Each pair of points delimit a region of PL defined by a rectangle (Fig. 2b). The size of a rectangular box is defined as having a width equal to the lateral distance between two points and a height equal to the axial space between the two points plus 3.08 mm. This added value corresponds to twice the pulse length (≈1.54 mm), which defines the axial resolution of the system. This value is included to account potential errors in the selection of the Npoints from the operator. The height of a rectangle designed between a pair of points extends from the deeper point plus 1.54 mm to the shallower point minus 1.54 mm (see Fig. 2b). As an example, a rectangular box designed between a pair points positioned at the depth of 100 and 92 mm will extend from 90.46 mm to 101.54 mm (with a height equal to 100−92+3.08 = 11.08 mm). The region of interest (ROI) fully containing PL is obtained by merging all the $(Npoints - 1)$ rectangles. However, this ROI is not optimal since other image components may be included (e.g., intercostal tissues and artifactual patterns). To remove unwanted components, the following strategy is used. Considering that the maximum intensity reflection is obtained from the PL, the image contained in the ROI is analyzed column-by-column along the depth dimension. The most intense point (MIP) depth position of each ROI column is used to define the PL. At the end of this process, given a ROI of L columns, PL is composed of L points (Fig. 2c). Afterwards, a margin is assigned to fully include PL. In this step it is important to specify that this margin does not represent the thickness of the physical pleural layers, composed by parietal pleura, pleural cavity, and visceral pleura. Instead, this layer corresponds to the thickness of the image representation of the acoustic interface between the intercostal-tissue and the air. While the thickness of the parietal and visceral pleura may be altered with different diseases, the acoustic interface thickness remains dependent on the axial resolution of the system.

2.3. Automatic segmentation algorithm of the pleural line (ASAP)

2.3.1. Algorithm pipeline

The amount of time required to manually segment PL is not negligible. This limit can be overcomed with the development of an automatic algorithm (AIM1). Although the manual segmentation of PL is realized with a AMSA algorithm, this approach continues to present a margin of subjectivity. Moreover, the time required to segment image-by-image is not negligible (using AMSA ≈ 0.10 images/s). These flaws can be overcomed with the development of ASAP. The aim of this algorithm is to automatically detect and segment the PL. Therefore, the following steps are deployed to compose ASAP.

Masking. ASAP starts by isolating the region containing PL from other LUS imaging patterns. This procedure is performed by masking the original images with a weighted ASAP mask (Fig. 3. A.I.a). The weighted ASAP mask is obtained as a column-by-column product between two components. The first component is an original 2D structural similarity (SSIM) mask (Fig. 3. A.I.b), designed for each acquisition. In detail, all the images of an acquisition centered at 3 MHz are selected and discretized from 0 to 255. To discretize the images, all the pixels with an intensity lower than −35 dB are assigned a value of −35 dB. (Fig. 3. A.I.a). The discretized images obtained are used only to compute the SSIM mask. This mask is computed image-by-image, starting from image 1 to image 23. For a group of 23 images, 22 masks of depth and lateral size of 4019 by 129 pixels are obtained, respectively. The masks are averaged and normalized with respect to the maximum value. These SSIM masks estimates the perceived changes in structural information within an acquisition. The application of SSIM is justified by PL movements caused by the respiratory cycle [40], in contrast to body tissues that are almost stationary. This will thus result in a low similarity in the areas where PL is present. The SSIM is used as follow $|SSIM - 1|$. The second component is a set of weights computed as a 1D Rayleigh cumulative density function (RCDF) (Fig. 3. A.I.c). RCDF is a function of the μ parameter (Fig. 3. A.I.d), which is in turn obtained as a product between two components. The first component is the average

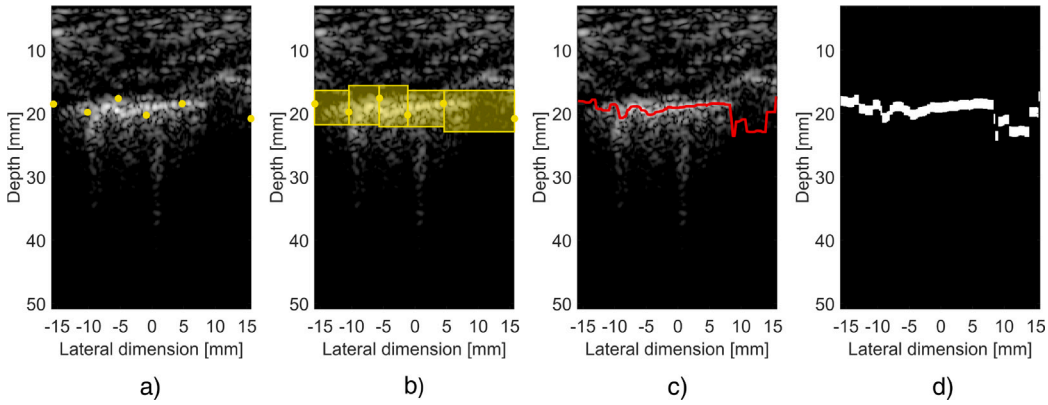

Fig. 2. Pipeline of the AMSA algorithm to detect the PL. (a) Manual part of the algorithm, where a series of points (yellow) defines PL position. (b), (c), and (d) Automatic part of the algorithm. In (b) a series of rectangles (yellow) that define the ROI are designed. (c) PL definition (red line) by means of the maximum intensity pixel for each column of the ROI defined in (b). (d) Definition of the binary mask (white) with the integration of ±50 points respect to the line defined in (c). The first three images are displayed in logarithmic scale with a dynamic range of 35-dB.

depth position of the most intense pixels (MIP) of the original image. The second component is a scale factor obtained from three FPs. If the area before the depth position of MIP is less than a free parameter (FP) named FP1, the scale factor is FP2, FP3 otherwise. The weighted SSIM mask, with 1 values representing moving components and 0 values representing more static components, is applied to the original outputs to isolates PL (Fig. 3. A.I.e)

Edges cancelling. From the isolated PL obtained in block (I), MIPs of each column are selected to generate a 1D first approximation of PL (Fig. 3. A.II.f). The shape of this line may be unrealistic compared to the anatomical shape of a real PL. Indeed, while PL may be partially visible in the image, the approximated PL is forced to select a point for all the 129 image lines. Consequently, this approximated PL may be composed by some anatomically unrealistic points (AUPs) (Fig. 3. A.II.f). To define a realistic representation of the PL, AUPs need to be removed. This is done through two different threshold procedures. Firstly, an intensity threshold is applied. The intensity threshold is computed as a difference among mean and variance of the approximated PL intensity (Fig. 3. A.II.g). The threshold is then applied to the approximated PL. Secondly, a spatial threshold is applied (Fig. 3. A.II.h). This threshold is obtained by computing the derivative depth position [$\Delta D(p)$] among pairs of consecutive points previously obtained. The number of $\Delta D(p)$ points exceeding FP4 are counted. If only one point exceed that value, the lateral dimension of ΔD position and the position of the highest MPI are compared. If the position of the highest MPI is most-left than ΔD, then the line will start from the beginning to ΔD; from ΔD to the end, otherwise. If more than one $\Delta D(p)$ point exceed 8 mm, then a spatial filter along the depth dimension is applied. This filter is obtained by grouping the series of points into FP5 depth ranges (Fig. 3. A.II.h). The depth range presenting the highest number of occurrences is selected and FP6 mm are subtracted. The number obtained represent the spatial threshold, under which PL points are removed (Fig. 3. A.II.i). FP values are finally obtained from the optimization process presented in Section 2.3.2.

Mask generation. In this final block, a 1D line is drawn by interpolating all the points obtained in block II with the MATLAB function *'interp1'*. The line is successively smoothed with the function *'smoothdata'*, using the method *'movmedian'* with window set to 5 (Fig. 3. A.II.j). Finally, a 2D mask that fully includes PL is defined. The 2D mask is designed by applying to the smoothed line a thickness equal to the pulse length in pixels. A total of ±50 points in the depth dimension of the line are added to define the PL mask (Fig. 3A.II.g).

Table 2
List of the six free parameters obtained for each outer k-fold and their corresponding value after optimization.

FP	Name	Value		
		Fold 1	Fold 2	Fold 3
FP1	Tissues Intensity	−35	−35	−35
FP2	Scale factor 1	1.75	2	2
FP3	Scale factor 2	0.1538	0.1538	0.1538
FP4	D(p) threshold	2	2	2
FP5	Depth ranges	229	261	269
FP6	Depth offset	202	202	206

2.3.2. Optimization

ASAP was optimized with a nested k-fold cross validation strategy on SAURON. Specifically, the dataset was cross validated with an outer k-fold partition, with k equal to 3, to split the dataset into training and test set. The training set was used to optimize the FPs by applying an inner k-fold strategy, with k equal to 3. For each inner k-fold, six FP values are obtained by means of the MATLAB function *"patternsearch"*. This derivative-free optimization method aims at finding a global minimum by analyzing different combinations of input parameters. The optimization is set by defining the lower and upper boundaries, as well as the starting points of the six parameters. The optimization options are set as follows: *"UseParallel" = True*, *"UseCompletePoll = True*, *"UseCompleteSearch" = True*, and *"AccelerateMesh = True*. The metric to be minimized is the Dice Loss function, computed among the masks output from ASAP and the masks from AMSA. The set of hyperparameters presenting the best Dice Loss in each inner k-fold is selected and used on the test set obtained from the outer k-fold. At the end of outer 3-folds, three Dice Loss values are obtained and averaged. The algorithm converged towards 55% of Dice loss (AIM2) with a stop criteria set when an improvement lower then 1e−4 was achieved. The set of six free parameters and their final value are reported in Table 2.

2.4. Features extraction

In the proposed approaches (Sections 2.2 and 2.3), PL position is defined with a binary mask. To achieve AIM3, the masks are used to segment the image into three regions: pre-pleura, pleura, and post-pleura. The regions are extracted for each multifrequency image as follows. The pre-pleura is defined as the region shallower than the segmented mask, the pleura is defined as the region contained in the mask, and the post-pleura is defined as the region situated deeper than the mask. The three regions have the same width, which is equal to the width of the mask. Indeed, the columns of the image not covered

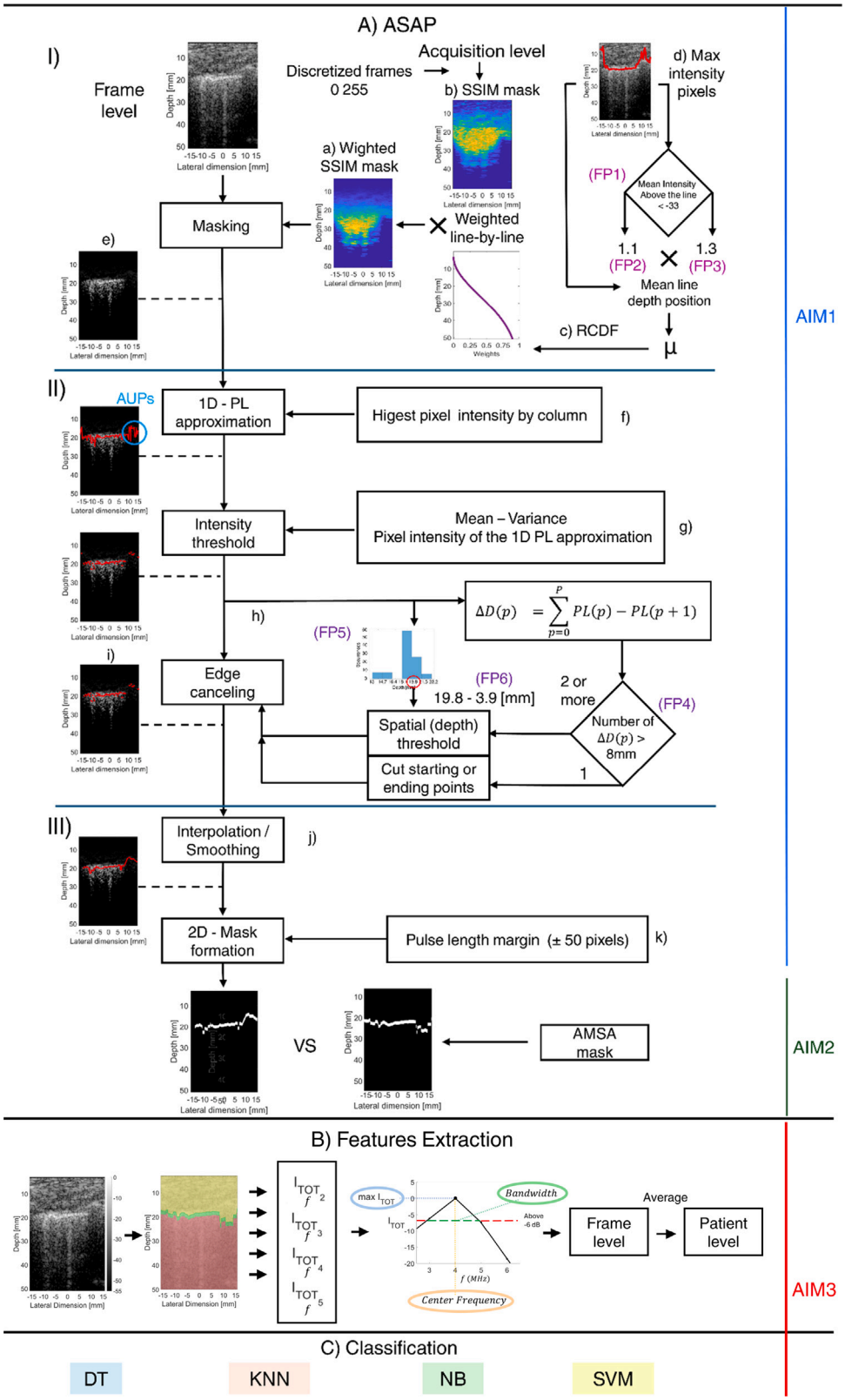

Fig. 3. Pipeline of the study, respectively: (A) ASAP pipeline, (I), masking block. The original image is masked with the weighted SSIM mask. The weighted SSIM mask represented in the image as a product between lines of the original SSIM mask and a Rayleigh CDF obtained as a function of μ.(II), the edges cancelling block describes intensity and spatial thresholds applied to remove unwanted points of PL, (III), the masking generation process, where to the new smoothed and interpolated 1D PL is applied a thickness equal to ±50 pixels. In purple, the six free parameters are highlighted, three in block (I) and three in block (II). (B) area parameterization and features extraction. The image is parameterized into 4 I_{TOT_n} n = 1, 2, 3, and 4. 3 features are extracted for each I_{TOT_n} to form the features at the frame level. Successively, each feature is average to represent each patient with 12 values. (C) The discrimination potential of the features is analyzed by means of 4 elementar classifiers: DT, KNN, NB, and SVM.

by the mask are neglected from the analysis. The segmented regions are parameterized by their I_{TOT}. For a multifrequency image, four values of pre-pleura, pleura, and post-pleura [$I_{TOT_1}(f)$, $I_{TOT_2}(f)$, and $I_{TOT_3}(f)$ (with f = 3, 4, 5, and 6 MHz)] are obtained, respectively. This parameter is defined as follows:

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

A_{PIX} is the pixel area ($\approx 3.7877 \times 10^{-3}$ mm^2 for both LUSARD and SAURON), i, j and R are the ith row and jth column of the pixel (with intensity greater than −35) in the R region (R = 1, 2, or 3 for pre-pleura, pleura, and post pleura, respectively) [6,29,31,35,36]. $ROI_{(i,j,R)}$ is the pixel intensity in logarithmic scale. In addition to these three parameters, two additional quartets were extracted: $I_{TOT_4}(f)$, defined as the ratio between $I_{TOT_1}(f)$ and $I_{TOT_2}(f)$, and $I_{TOT_5}(f)$, defined as the ratio between $I_{TOT_2}(f)$ and $I_{TOT_3}(f)$. These two ratios describe the relationship between image intensity over depth and PL reflectivity variations. Since the values are in logarithmic scale, the ratio is computed as a difference between the two values. The ultrasound spectroscopy is performed by extracting from each parameter: center frequency, bandwidth, and $\max_f I_{TOT_R}$ [29,31,36], where $\max_f I_{TOT_R}$ ($R = 1, 2, \ldots, 5$) represents the maximum $I_{TOT_R}(f)$ as a function of frequency. The center frequency is defined as the frequency at which the $\max_f I_{TOT}$ is obtained (Fig. 3.B). The bandwidth is defined as the range of frequencies having a normalized intensity greater than −6 dB with respect to the maximum value [31]. All the features were extracted from 4 parameters, with I_{TOT_1} being excluded as it does not include ultrasound signal backscattered from the lung. At the end, for each patient, we will have 12 values (3 features × 4 parameters). To obtain a value for each feature for a given patient, the average with respect to all her/his images were computed (Fig. 3.B).

2.5. Classification for differential diagnosis

The differential diagnosis performance of the features extracted from the I_{TOT_R} (Fig. 3.C) parameters are tested through decision tree (DT) (with *MinParentSize* set to 10), Support Vector Machine (SVM), k-nearest neighbors (KNN), and Naive Bayes (NB) by the MATLAB functions *fitctree, fitcsvm, fitcknn,* and *fitcnb*, respectively.

2.5.1. Binary classification

The significance of the extracted features is tested by computing the p-value through a Welch's t-test with the MATLAB function *"ttest2"*. The p-value was computed for all the features across different combination of diseases.

Afterwards, four different binary classifiers are tested (Fig. 3.c) on LUSARD data, as it is the dataset to which we can benchmark [31]. Each type of classifier was trained with 7 input combinations [obtained from the combination of the three features of an I_{TOT_R} parameter; combination with single feature (3) + combination with features pairs (3) + combination of 3 features (1)], resulting in a total of 112 classifiers (7 combinations × 4 I_{TOT_R} × 4 classifiers). To differentiate PNE-CPE, 112 classifiers were trained on LUSARD data. To differentiate COPD-PF, COPD-PFE, and PF-PFE, 112 classifiers for each couple of diseases were trained on SAURON data. The classifiers were trained with a split at the patient-level to avoid leakage from correlated images of the same subject. All the classifiers were trained with a repeated k-fold strategy, where k ranges from 2 to 13. Additionally, possible biases caused by the random choice of the samples in the k-fold strategy are mitigated by repeating each k-value 20 times. The performance of each classifier is evaluated by means of accuracy, specificity, and sensitivity. Specifically, the results of each k-value are obtained by averaging the 20 repeated validations.

Each classifier performance are then described by selecting the minimum, mean, and maximum values of accuracy, specificity, and sensitivity for each set of k-values. The results obtained are then compared with the state of the art [31].

Additionally to AIM3, we tested the differential diagnosis performance also on SAURON dataset, showing different classification performance among ASAP and AMSA segmentation approaches. The same binary classifier and the same validation strategy were utilized (except for the k-fold, which ranges between 2 and 8) for pair of diseases (3 combination of pair of diseases, given that SAURON presents 3 diseases). Results demonstrated improved performance of the proposed approach with respect to state of the art and expand the analysis to different pathologies analyzed here for the first time.

2.5.2. Multiclass classification

To test the approach capability under a clinically realistic scenario, a multiclass classification with the four standard classifiers was performed. Patients from LUSARD and SAURON, along with their spectral features were aggregated, forming a single dataset of 85 patients with twelve spectral features and affected by five different lung diseases. To select the best features we utilized two different approaches. In the first approach, all the possible $2^{12} - 1$ combinations of input features were analyzed [i.e., combination with single feature (12) + combination with features pairs (66) + combination of 3 features (220) + combination of 4 features (495)... + combination of 12 features (1)], resulting in a total of 4095 combinations of features for each standard classifier. The best accuracy obtained for each number of input features is selected, obtaining in total 12 different accuracies. In the second approach, a rank of the most significant features in differentiating the five lung diseases was obtained by means of the ReliefF algorithm (implemented with the MATLAB function "relieff") [41]. Afterwards, an incremental feature addition test was performed, analyzing the performance of a classifier by adding different features in order of significance (e.g., 1th classifier is trained with the most significant feature, 2th classifier with the first two most significant features, 3° classifier with the first three most significant features... 12th classifier is trained with all the 12 features). The multiclass classifiers were trained following a repeated K-fold cross validation strategy, with K equal to 8 and repeated 10 times. The partitioning of the training samples was set to have an equal partition among the five diseases.

Table 3
Evaluation metrics between ASAP (generated data) and AMSA (ground-truth) masks.

Metric	Score
Intersection over union	35.4%
Dice	52.2%
Accuracy	97.7%
Structure similarity	95.2%
Boundary F-measures	70.3%

3. Results

3.1. Segmentation

Aim 1 and 2 results are reported in this paragraph. ASAP algorithm was optimized and tested by means of a nested k-fold cross validation. Three Dice loss values are obtained from the test of the outer k-fold, utilizing the set of hyperparameter optimized in the inner k-fold (Table 2). These three Dice loss values are averaged, obtaining a Dice Loss of 55%.

Table 3 reports the evaluation metrics obtained from ASAP (on SAURON) with respect to the ground truth. Intersection over union (IoU), Dice score, structure similarity, and the boundary F-measures were obtained by means of the MATLAB functions: *"jaccard"*, *"dice"*, *"ssim"*, and *"bfscore"*, respectively. Accuracy was computed as the ratio in percentage of the sum of the pixels with the same value among the two masks, and the number of all the pixels of an image (129 × 4096).

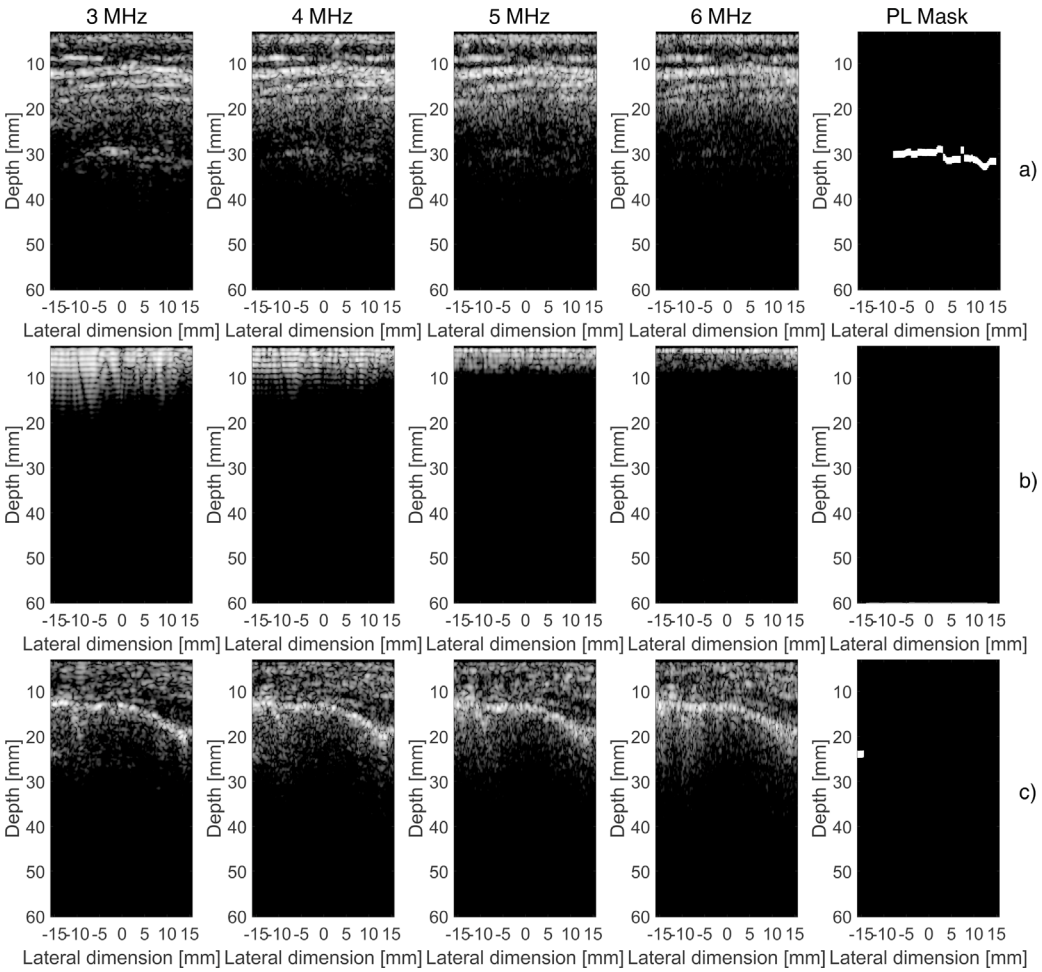

Fig. 4. Example of three segmentations excluded from the spectroscopy analysis. Each line represent an example. Example (a), the absence of PL at 3 MHz of center frequency. Example (b) absence of LUS images. Example (c), ASAP error. The first four columns of the image depicts the center frequency of the multifreqeuncy frame, the last column depicts the mask extracted from ASAP.

LUSARD dataset was segmented with ASAP in ≈ 300 s (≈ 10 images/s). In six images, non-utilizable I_{TOT} values were found and were excluded from the differential analysis (rejection rate of 0.22%).

Fig. 4 illustrates three different reasons for the exclusion of LUS images from the classification phase. The first reason is the absence of PL at one of the center frequencies. In the first row, the first four images representing a multifrequency image show PL at 3, 4, and 5 MHz; however, at 6 MHz, PL is absent, leading to an $\max_4 I_{TOT} = -\infty$. Consequently, the image is excluded from the spectral analysis. The second reason is the absence of identifiable LUS imaging patterns. In the second row, these images are removed from the spectroscopy analysis since they do not display any type of LUS imaging pattern, probably due to an erroneous probe placement or an excessive patient movement during the acquisition phase. The last reason is due to an erroneous segmentation of PL from ASAP. This error is related to the shape of PL and its effect on the thresholding system during the edge cancelling phase (see Fig. 3.A.II). Indeed, in this scenario the algorithm is unable to follow PL shape, incorrectly thresholding PL and cutting it off at the wrong depth.

3.2. Classification

3.2.1. Binary classification

Table 4 depicts the p-values obtained from the statistical significance test. For the couples PNE-CPE and PF-PFE the test shown no significance for the extracted features. Contrary, a couple of diseases such as COPD-PF and COPD-PFE presents multiple significant features.

The results of AIM3 are summarized in Fig. 7(a), showing the classification performance of PNE vs CPE. Each subplot is divided into four colors to represent the type of classifier used among DT (blue), KNN (red), NB (green), and SVM (yellow). The performance of each classifier are represented by means of accuracy (blue), sensitivity (red), and specificity (green). The best accuracy achieved by each classifier is highlighted in orange. PNE vs CPE shows classification performance of 77.28% of accuracy when considering the bandwidth and $\max_f I_{TOT}$ of I_{TOT_5}, with sensitivity and specificity values of 73.43 and 80.61%, respectively. The proposed approach outperforms the state of the art in terms of accuracy and sensitivity [31].

The results of the additional test conducted on SAURON are depicted in Fig. 7(b,c,d), showing the differential diagnosis of COPD vs PF (a), COPD vs PFE (b), PF vs PFE (c). All these analysis are based on segmentations obtained by ASAP. COPD vs PF presents an accuracy of 78.61% with the SVM classifier, when considering the bandwidth (B) of I_{TOT_3}, with sensitivity and specificity values of 72.26 and 84.64%, respectively. COPD vs PFE reaches accuracy values up to 87.48% with the KNN classifier, when considering the center frequency (A) of I_{TOT_2}, with sensitivity and specificity values of 80.42 and 87.48%, respectively. PF vs PFE reaches accuracy values up to 74.02% with the DT classifier, when considering $\max_f I_{TOT}$ (C) of I_{TOT_5}, with sensitivity

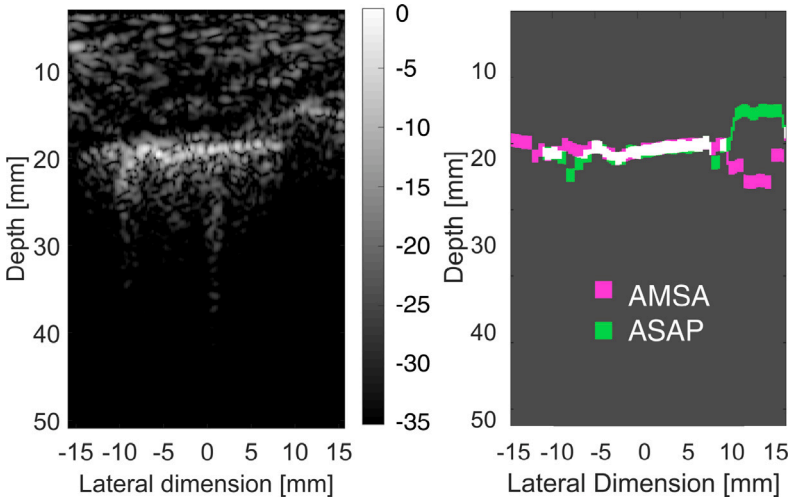

Fig. 5. Example of segmentations (right) performed on the same LUS image (left) from AMSA (pink) and ASAP (green).

Table 4
p-values of all the features among different diseases. The significant values are highlighted in green.

		PNE CPE	COPD PF	COPD PFE	PF PFE
Center Frequency	I_{TOT2}	7.5e−1	1.9e−3	8.6e−5	5.2e−1
	I_{TOT3}	9.7e−1	3.7e−2	4.5e−2	7.2e−1
	I_{TOT4}	7.2e−1	1.2e−2	5.9e−3	4.5e−1
	I_{TOT5}	7.2e−1	3.8e−2	4.9e−1	6.0e−1
Bandwidth	I_{TOT2}	4.4e−1	6.2e−2	1.5e−2	8.0e−1
	I_{TOT3}	2.5e−1	7.2e−4	6.6e−3	9.4e−1
	I_{TOT4}	5.6e−1	8.5e−2	1.2e−2	3.6e−1
	I_{TOT5}	2.0e−1	1.7e−1	3.0e−2	2.2e−1
$\max_f I_{TOT}$	I_{TOT2}	6.1e−1	4.1e−1	6.7e−1	8.2e−1
	I_{TOT3}	5.5e−1	2.4e−1	5.7e−3	4.9e−2
	I_{TOT4}	6.3e−1	8.5e−1	3.6e−2	7.9e−2
	I_{TOT5}	1.1e−1	3.4e−3	6.2e−5	1.5e−2

and specificity values of 55.18 and 85.32%, respectively. These results show, for the first time, the potential of the proposed technique to diagnose COPD and PF and, within PF, to stratify between PF and PFE.

In addition, Fig. 6 depicts the comparison on the classification performance when segmenting with ASAP and AMSA (SAURON), respectively. Each background color suggests the type of classifier used among DT (blue), KNN (red), NB (green), and SVM (yellow). Each sub-figure is divided into three parts, each representing a pair of diseases (COPD vs PF, COPD vs PFE, and PF vs PFE from left to right, respectively). The classification performance of ASAP with accuracy above 70% are highlighted with blue rectangles. The performance highlighted in yellow represents the best accuracy values above 70% for each I_{TOT_R}. The heatmaps represent the difference in accuracy between ASAP and AMSA. Precisely, a negative value on the graph (red scale) indicates that ASAP approach performs worse than AMSA. In contrast, a positive value (green scale) indicates that ASAP outperforms AMSA approach. For example, the only blue rectangle highlighted from the classification of PF vs PFE with the KNN (red background) indicates that ASAP exceeds the 70% of accuracy. This performance is obtained using $\max_f I_{TOT}$ (feature C) of I_{TOT_5}. The cell is also highlighted in yellow, indicating the best performance for that classifier. The cell shows a positive green value of +3.88, indicating a performance improvement of +3.88% for ASAP compared to the AMSA algorithm.

In Fig. 6d, the SVM classifier presents more similar accuracy values between the two segmentation approaches, ranging between −15.0 and 12.9%. In contrast, the NB classifier (Fig. 6c), presents larger differences with values ranging between −23.6 and 23.2%. Among the diseases, COPD vs PFE presents larger differences between the two segmentation approaches, ranging between −23.6 and 23.2%. COPD vs PF instead ranges between −16.2 and 12.5% and PF vs PFE between −1.79 and 10.4% for all the classifiers. The highest number of cases presenting an accuracy above 70% (highlighted with blue rectangles) is achieved for COPD vs PFE (6, 5, 16, 14 for DT, KNN, NB, and SVM, respectively). PF vs PFE instead presents the lowest number of cases with accuracy above 70% (3, 1, 0, and 4 for DT, KNN, NB, and SVM, respectively). For COPD vs PF, the highest number of cases presenting accuracy values above 70% is achieved using DT. The results depicts how, in some cases, ASAP outperform AMSA. When AMSA outperform ASAP, ASAP maintains diagnostic performance with accuracies above 70%.

3.2.2. Multiclass classification

Fig. 8 highlights the results of the multiclass classifiers in differentiating between the five lung diseases. Fig. 8a depicts the best results achieved from each number of input features through the $2^{12}-1$ combinations of input features. The best result is achieved utilizing the KNN classifiers with an accuracy of 65%. This result is obtained utilizing the Center Frequency of I_{TOT2}, I_{TOT4}, BW I_{TOT2}, I_{TOT3}, and I_{TOT5}. The figure shows how the optimal number of features ranges from 4 to 8; beyond this range the classifiers tend to overfit. Fig. 8b shows the best classification performance of the multiclass classifier trained through an incremental feature addition test. The classifiers show the best accuracy of 58% with the combination of 8 features, when the SVM is utilized. The final results show the best classification performance with the features obtained from by the feature combination approach, achieving accuracies of 65% with KNN, against 58% of accuracy with SVM obtained with the incremental feature addition test.

3.3. Latency test

To test the real-time potential of the proposed methodology, an end-to-end latency test was performed. We simulate a clinical scenario of a patient with 12 body areas scanned, each scan composed by 23 multifrequency images. Our proposed methodology proved able to segment all the 276 images, extract the features, and classify the patient

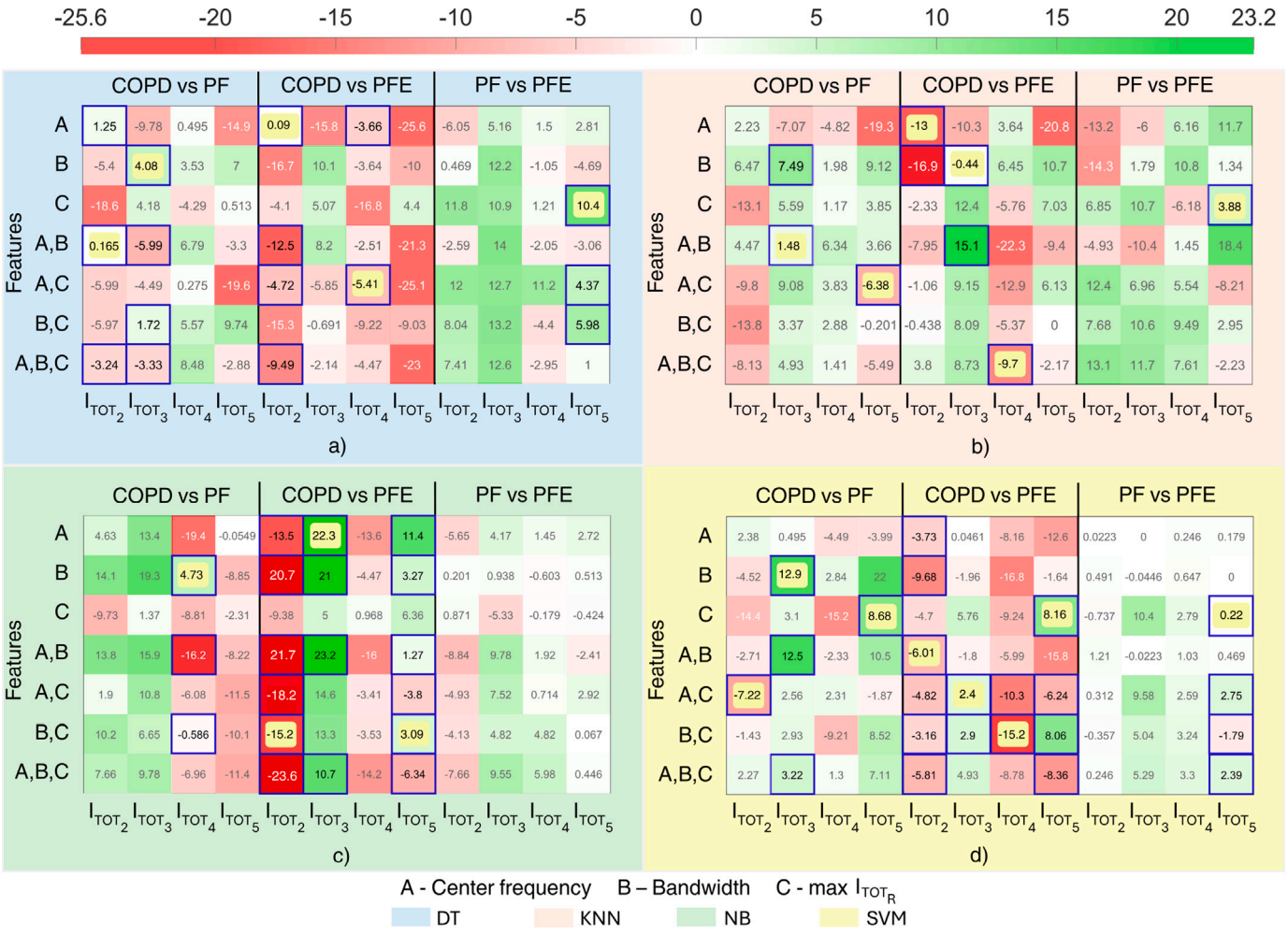

Fig. 6. Heatmaps representing the difference of accuracies expressed in percentage between the classifiers trained using ASAP and the AMSA (SAURON). The figure shows four heatmaps, distinguished by type of classifier: (a) DT, (b) KNN, (c) NB, and (d) SVM. Each heatmap is grouped into four columns to represent the classification between COPD vs PF, COPD vs PFE, and PF vs PFE. Each quadruple depicts I_{TOT_2}, I_{TOT_3}, I_{TOT_4}, and I_{TOT_5}. The *y*-axis represents the features used to train the classifier, where A is the center frequency, B the bandwidth, and C is the max I_{TOT_R}. The colors of the heatmaps range between −25.6 to 23.2. The cells highlighted in blue represent the accuracies above 70% reached by ASAP. The cells highlighted in yellow represent the best I_{TOT_R} accuracy above 70%.

in 19.2 s. This result corresponds an amount of 14.4 images processed per second. The test was run on MATLAB utilizing a 13th Gen Intel(R) Core(TM) i7-13700T (24 CPUs) with 16 GB RAM.

4. Discussion and conclusion

LUS is an imaging tool used to assess the condition of the lung surface in real-time. However, LUS is a qualitative and subjective practice that leads to poor diagnostic specificity [1,26]. Novel studies have focused on the quantification of LUS patterns to improve specificity [31,42]. Nonetheless, these approaches were still influenced by the subjective and time-consuming detection of LUS imaging patterns [31].

This article addressed three main aims: aim 1, we developed ASAP algorithm to automatically segment PLs of RF images in quasi real-time. In this process we developed AMSA to manually segment SAURON as a ground-truth and optimize ASAP.

In aim 2, we validated ASAP with respect to AMSA. In terms of segmentation rate, ASAP segmented 10,024 images in less than 1000 s, achieving a rate of 10 images/s and ensuring quasi real-time performance. AMSA presented an average rate of ≈ 0.10 images/s, highlighting how ASAP segments ≈100 time faster than AMSA. ASAP accuracy, Dice score, structure similarity, intersection over union, and boundary F-measures are respectively 97.7%, 51.0%, 95.2%, 35.4%, and 70.3%. The Dice score is computed on masks with maximum size 101 × 129 in depth and lateral dimension, respectively, corresponding to ≈ 2.5% of a whole image.

It is important to clarify that ASAP presented 68 images over 10,024 (less than 0.68% of the whole dataset) in which the quantification of PL gave an $\max_4 I_{TOT} = -\infty$, meaning PL mask was defined in a region not presenting values above −35 dB [29,31]. Given the low percentage of images that were discarded, it is plausible to expect a negligible impact of those images in the final classification. However, these images were considered for the computation of the evaluation metrics of aim 2. The 10,024 images were automatically segmented in less than ≈1000 s (≈10 images/s).

The masks obtained from both ASAP and AMSA are used to split the LUS images into three regions: pre-pleura, pleura, and post-pleura. These regions are quantified into five I_{TOT_R} $(R = 1, 2, 3, 4, \text{ and } 5)$ parameters and three spectral features are extracted from each parameter: center frequency, bandwidth, and $\max_f I_{TOT}$ [29,31,36].

In aim 3, we applied a novel quantitative LUS approach to differentiate lung diseases on LUSARD and we compared the results with the state of the art. This dataset is composed of patients affected by PNE and CPE. These diseases were of particular interest for their histopathology and social impact. Indeed, CPE is a condition that occurs with heart failure and is among the leading factors of hospitalization

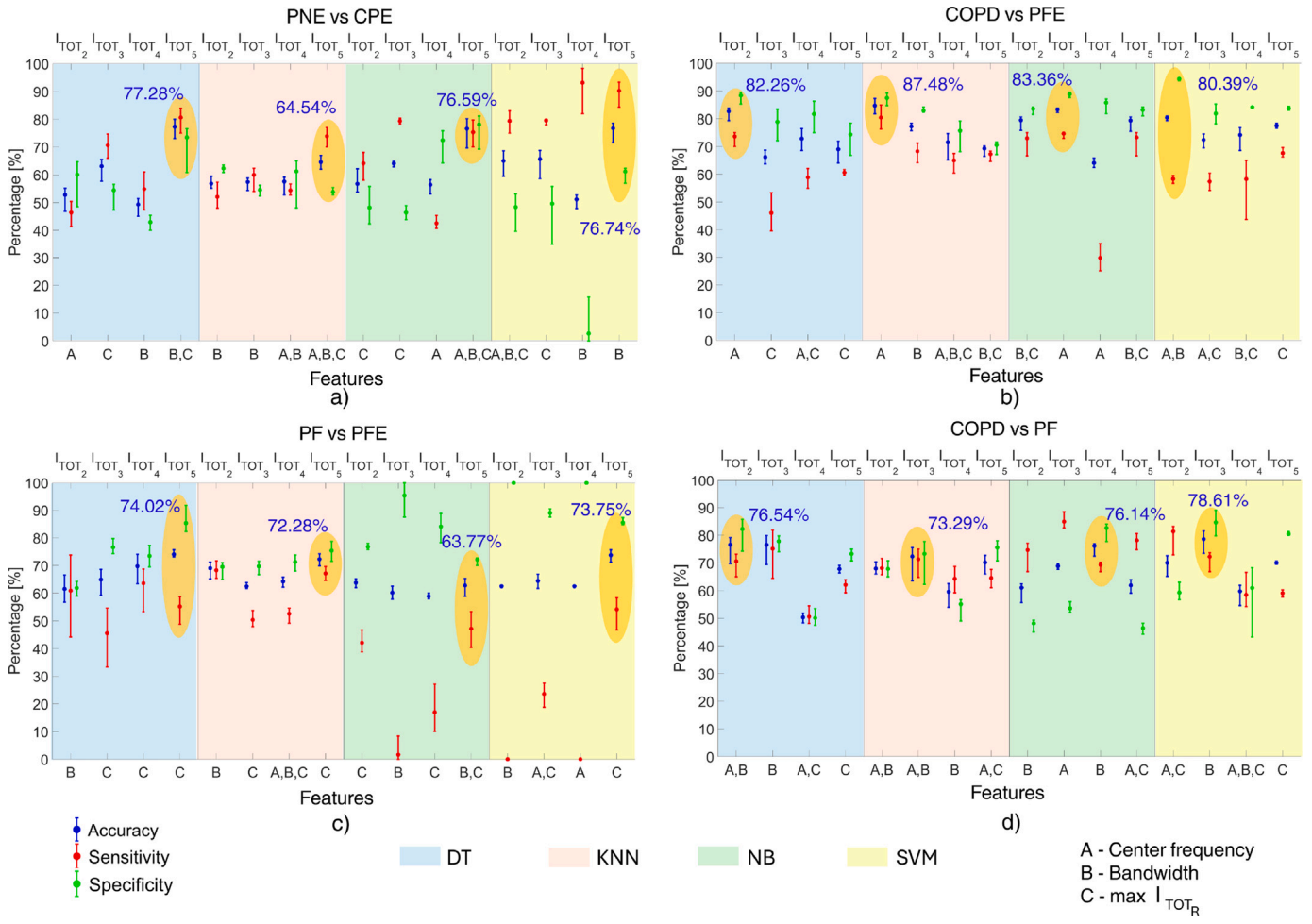

Fig. 7. Four graphs (errorbars) depict the discrimination results in terms of accuracy (blue bars), sensitivity (red bars), and specificity (green bars). Each graph show the discrimination performance between two diseases: COPD vs PF (a), COPD vs PFE (b) and PF vs PFE (c) from SAURON dataset; PNE vs CPE (d) from LUSARD dataset. Each graph is divided into four colored regions representing a different type of classifier: BT (light blue), KNN (light red), NB (light green), and SVM (light yellow). For each classifier, the best result obtain of I_{TOT_2}, I_{TOT_3}, I_{TOT_4}, and I_{TOT_5} are reported (label on the top x-axis) along with the features utilized (bottom x-axis) [(A - center frequency, B - bandwidth, and C - $\max_f I_{TOT}$. Multiple features are represented with a comma (i.e., A,B - center frequency and bandwidth)]. The y-axis represents the percentage obtained between 0 to 100%. The best results obtained with each classifier are highlighted with an orange elipse and the corresponding value.

among people over 65 [43,44]. PNE is the most common infection causing death among children [45].

LUSARD is automatically segmented with ASAP, and center frequency, bandwidth, and $\max_f I_{TOT}$ were obtained. Thanks to ASAP optimization performed only on SAURON, the feature values obtained by segmenting LUSARD with ASAP demonstrate a binary classification improvement with respect to the state of the art.

The features are tested through four different binary classifiers and the results are represented in Fig. 7(a). The best classification performance in terms of accuracy, specificity, and sensitivity are obtained with KNN and DT. To compare the results of the proposed approach with the state of the art, the vertical artifacts of PNE and CPE patients are also analyzed [31]. The classification analyses for both approaches were performed using the same cross-validation split folds. The vertical artifacts approach achieves an accuracy of 71.00% in classifying these two diseases. The proposed approach outperformed the state of the art by more than 6%, achieving an accuracy of 77.28%, with high levels of sensitivity (81%) and specificity (73%). Additionally, our approach is based on the detection of anatomical patterns, which, contrary to vertical artifacts, are less affected by variation of imaging settings.

The additional analysis performed consisted of applying for the first time the differential diagnosis to SAURON, with three new lung diseases: PF, PFE, and COPD. PF and PFE represent different stages of the same pathology. They both are chronic lung conditions characterized by scarring of lung tissues leading to poor prognosis. PFE is a worsening condition of PF [29,32,42,46]. COPD is the fourth leading cause of death worldwide, causing 3.5 million deaths in 2021, approximately 5% of all global deaths [47]. The binary classification results obtained from ASAP segmentation are illustrated in Fig. 7(b,c,d). The best classification performance are obtained with KNN and NB classifiers, showing accuracies up to 87.48% in discriminating COPD vs PFE, up to 78.61% in discriminating COPD vs PF, and up to 74.02% in discriminating PF vs PFE.

The results of the Welch's t-test enforce the results of the classifiers, although it was performed over a limited group of patients (most populated group is PF with 20 patients), affecting the outcome of the test. Additional data are needed to improve the robustness of the test. PNE-CPE and PF-PFE respectively show a low number of significant features, which may be related to similarities in their histopathology. Respectively, PNE-CPE shows classification accuracies up to 77.28% when utilizing the combination of bandwidth and $\max_f I_{TOT}$ of I_{TOT5}, which represent the two features with the lowest P-value (even though p-value > 0.05), while PF-PFE shows accuracies up to 74.02% with the bandwidth of I_{TOT5}, which represents the feature with the most significant p-value for this group of diseases (0.0015).

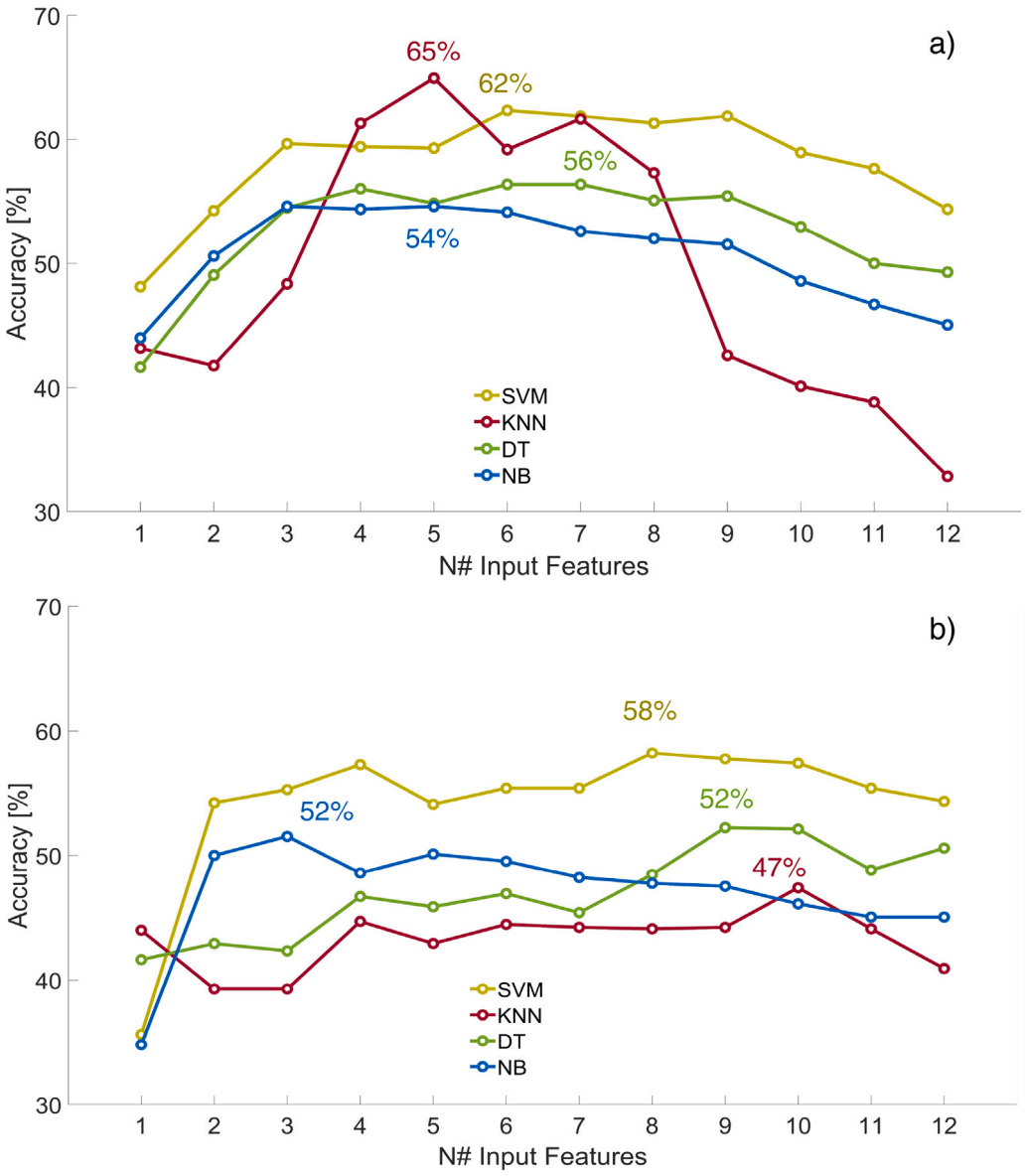

Fig. 8. multiclass classifiers representing the best accuracies achieved by number of input feature. In (a), the best input features are selected by analyzing all the $2^{12}-1$ possible combinations of input features. In (b), the input features are selected by the ReliefF algorithm. In the *y*-axis is represented the accuracy, while in the *x*-axis the number of input features utilized. The graph shows the results of SVM (yellow), KNN (red), DT (green), and NB (blue).

Results depict how this approach can support the clinical diagnosis of PF and PFE vs COPD, and stratify two different conditions of the same pathology (PF and PFE). Additionally, these results potentially confirm the efficiency of KNN with respect to other classifiers in differentiating lung diseases with these spectroscopy features [31,36,42].

A comparison of the differential analysis performance between ASAP and AMSA are represented in Fig. 6. The graphs show the differences in terms of accuracies, highlighting the results above 70% reached from ASAP only. Notably, COPD vs PFE present, for all the classifiers, the highest number of cases presenting accuracies above 70%. Accordingly, the histopathology of PFE presents a reduction of the air-spaces dimensions, while COPD an enlargement of the air-spaces dimensions. This leads to different PL frequency responses that can differentiate the two diseases [35]. PF vs PFE presents the lowest number of values presenting accuracies above 70%. We expect this result, as the two diseases present a similar histopathology.

Possible discrepancies between ASAP and AMSA are dictated by the different PL segmentation between operator and algorithm. Indeed, the segmentation of PL mask has a strong impact on the discrimination performance. The manual segmentation performed through AMSA algorithm could be biased by the interpretation of a video by the operator. Indeed, the operator has access to the video and the appearance of PL on different images could lead the operator to interpolate and predict the position of PL, even with the absence of PL (e.g., if from image 1 to 10 PL is visible at −15 mm along the lateral dimension, the operator may select PL in that part of frame 11, even if PL is not visible) (see Fig. 2). This aspect affects the selection of the initial and final points of PL, generating a wider PL mask that may include parts of the image with intensities below −35 dB. Similarly, in some images, PL may appear discontinuous, potentially presenting parts with intensities lower than −35 dB, which affect PL detection in that area. Even though a rectangular ROI may include a part of the image with intensities lower than −35 dB, AMSA remains based on the same pipeline. Specifically, AMSA selects the most intense pixel column-by-column, independently on the absolute value of the pixel. Notwithstanding the selection of the above-mentioned parts, the downstream analysis remains unchanged, given that all the points with intensity values lower than −35 dB are

later removed from the analysis. An alternative approach to generate ground truth would consist of a pixel-by-pixel manual segmentation, which would be more subjective and less reproducible compared to AMSA.

Contrary, ASAP is not affected by this type of bias, but is based on objective evaluations of the images (see Fig. 5. Since these evaluations are based on thresholding systems, in some cases the drawn masks could be less precise than the manual segmentation. Moreover, the maximum dimension of a PL mask corresponds to the 2.5% of the pixels composing an image. This highlights the complexity of this segmentation task. Differences in the segmentation metrics could be dictated by the low percentage of pixels composing the PL, as reported also from other articles [17]. In addition to these metrics, both AMSA and ASAP segmentations have been validated by three expert clinicians with over ten years of experience in LUS (authors AS, RI, TP), which confirmed the clinical validity of the segmentations. Indeed, the results of the final clinical task, consisting on the differential diagnosis of lung diseases, outperform the state of the art, implicitly confirming the clinical validity of the proposed approach.

For the first time in literature, a multiclass classification of five lung diseases was performed utilizing a quantitative LUS spectroscopy technique. The results exceeded the performance of the state of the art (obtained with only three diseases) [48] by 5%. Although the multiclass classifiers results could be biased by the acquisition phase of SAURON and LUSARD, the two datasets have been acquired with the same scanner and acquisition settings.

Notably, the latency test was performed on a code not designed at the maximum efficiency in term of latency. Different code implementation may improve the latency performance. Moreover, the test was performed on a patient with 12 acquired areas (maximum amount of areas following the standardized acquisition protocol) [49]. The final end-to-end latency test demonstrate the applicability of our proposed technique in a quasi real-time scenario.

In conclusion, in this study we provided a segmentation approach along with a new quantitative LUS approach to automatically segment in quasi real-time, quantify, and differentiate RF data of patients acquired from different centers and affected by different lung diseases. The results show how the proposed technique outperforms the state of the art in terms of segmentation time, differential diagnosis performance, and reproducibility. Additionally, a new set of lung diseases are analyzed for the first time: PFE and COPD.

Some limitations need to be highlighted. The acquisition strategy utilized on LUSARD may not reflect routine scans. LUSARD was previously designed to perform a spectroscopy analysis on vertical artifacts only [31]. Nevertheless, this dataset allowed us to compare the results of our newly developed approach with the state of the art.

The study focused on the analysis of two different linear probes of the same manufacture. Even though a previous study demonstrated how the linear probe outperforms convex probe due to its wider available bandwidth [31], future studies could investigate the performance of probes from different manufacturers, as long as they are able to cover the same bandwidth, thus allowing a comparable spectroscopy analysis. Additionally, the examination of same pathologies acquired from different centers can be explored to enforce the validity of this multicenter study. Moreover, the quantitative approach can be applied to lung diseases never explored with quantitative LUS.

CRediT authorship contribution statement

Mattia Perpenti: Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Federico Mento:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Giovanni Pierro:** Resources. **Alessandro Perrotta:** Resources. **Tiziano Perrone:** Resources. **Andrea Smargiassi:** Resources. **Riccardo Inchingolo:** Resources. **Libertario Demi:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

That generative AI and AI-assisted technologies have not been utilized in the writing process or if used, disclosed in the manuscript the use of AI and AI-assisted technologies and a statement will appear in the published work.

That generative AI and AI-assisted technologies have not been used to create or alter images unless specifically used as part of the research design where such use must be described in a reproducible manner in the methods section.

Funding

This work was partially supported by a research grant provided in compliance with Mission 6/component 2/Investment: 2.1 ”Strengthening and enhancing biomedical research in the NHS”, financed by the European Union - NextGenerationEU, CUP: E63C24000660007.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Libertario Demi, Frank Wolfram, Catherine Klersy, Annalisa De Silvestri, Virginia Ferretti, Marie Muller, Douglas Miller, Francesco Feletti, Marcin Weł nicki, Natalia Buda, Agnieszka Skoczylas, Andrzej Pomiecko, Domagoj Damjanovic, Robert Olszewski, Andrew Kirkpatrick, Raoul Breitkreutz, Gebhart Mathis, Gino Soldati, Andrea Smargiassi, Tiziano Perrone, New international guidelines and consensus on the use of lung ultrasound, J. Ultrasound Med. 42 (2022).

[2] Libertario Demi, Thomas Egan, Marie Muller, Lung ultrasound imaging, a technical review, Appl. Sci. 10 (2020) 462, http://dx.doi.org/10.3390/app10020462, URL https://www.mdpi.com/2076-3417/10/2/462.

[3] Libertario Demi, Lung ultrasound: The future ahead and the lessons learned from COVID-19, J. Acoust. Soc. Am. 148 (2020) 2146–2150, http://dx.doi.org/10.1121/10.0002183, URL https://pubs.aip.org/asa/jasa/article/148/4/2146/994360/Lung-ultrasound-The-future-ahead-and-the-lessons.

[4] M. Demi, R. Prediletto, G. Soldati, L. Demi, Physical mechanisms providing clinical information from ultrasound lung images: Hypotheses and early confirmations, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 67 (3) (2020) 612–623, http://dx.doi.org/10.1109/TUFFC.2019.2949597, URL https://ieeexplore.ieee.org/document/8883203.

[5] Gino Soldati, Marcello Demi, Libertario Demi, Ultrasound patterns of pulmonary edema, Ann. Transl. Med. 7 (Suppl 1) (2019) URL https://atm.amegroups.org/article/view/23881/23195.

[6] Federico Mento, Libertario Demi, On the influence of imaging parameters on lung ultrasound B-line artifacts, in vitro study, J. Acoust. Soc. Am. 148 (2020) 975–983, http://dx.doi.org/10.1121/10.0001797, URL https://pubs.aip.org/asa/jasa/article/148/2/975/960282/On-the-influence-of-imaging-parameters-on-lung.

[7] Federico Mento, Libertario Demi, Dependence of lung ultrasound vertical artifacts on frequency, bandwidth, focus and angle of incidence: An in vitro study, J. Acoust. Soc. Am. 150 (2021) 4075–4082, http://dx.doi.org/10.1121/10.0007482, URL https://pubs.aip.org/asa/jasa/article/150/6/4075/993698/Dependence-of-lung-ultrasound-vertical-artifacts.

[8] Berke Cansiz, Coskuvar Utkan Kilinc, Gorkem Serbes, Deep learning-driven feature engineering for lung disease classification through electrical impedance tomography imaging, Biomed. Signal Process. Control. (ISSN: 1746-8094) 100 (2025) 107124, http://dx.doi.org/10.1016/j.bspc.2024.107124, URL https://www.sciencedirect.com/science/article/pii/S1746809424011820.

[9] Berke Cansiz, Coskuvar Utkan Kilinc, Gorkem Serbes, Tunable Q-factor wavelet transform based lung signal decomposition and statistical feature extraction for effective lung disease classification, Comput. Biol. Med. (ISSN: 0010-4825) 178 (2024) 108698, http://dx.doi.org/10.1016/j.compbiomed.2024.108698, URL https://www.sciencedirect.com/science/article/pii/S0010482524007832.

[10] Hamza Ilhan, Gorkem Serbes, Nizamettin Aydin, Decision and feature level fusion of deep features extracted from public COVID-19 data-sets, Appl. Intell. 52 (2021) http://dx.doi.org/10.1007/s10489-021-02945-8.

[11] Sezer Ulukaya, Gorkem Serbes, Yasemin P. Kahya, Resonance based separation and energy based classification of lung sounds using tunable wavelet transform, Comput. Biol. Med. (ISSN: 0010-4825) 131 (2021) 104288, http://dx.doi.org/10.1016/j.compbiomed.2021.104288, URL https://www.sciencedirect.com/science/article/pii/S0010482521000822.

[12] Sezer Ulukaya, Gorkem Serbes, Yasemin P. Kahya, Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique, Comput. Biol. Med. (ISSN: 0010-4825) 104 (2019) 175–182, http://dx.doi.org/10.1016/j.compbiomed.2018.11.004, URL https://www.sciencedirect.com/science/article/pii/S0010482518303421.

[13] Federico Mento, Tiziano Perrone, Anna Fiengo, Andrea Smargiassi, Riccardo Inchingolo, Gino Soldati, Libertario Demi, Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter studya), J. Acoust. Soc. Am. (ISSN: 0001-4966) 149 (5) (2021) 3626–3634, http://dx.doi.org/10.1121/10.0004855.

[14] Tatiana Khokhlova, Gilles Thomas, Jane Hall, Kyle Steinbock, Jeff Thiel, Bryan Cunitz, Michael Bailey, Layla Anderson, Ross Kessler, M. Hall, Adeyinka Adedipe, Development of an automated ultrasound signal indicator of lung interstitial syndrome, J. Ultrasound Med. : Off. J. Am. Inst. Ultrasound Med. 43 (2023) http://dx.doi.org/10.1002/jum.16383, URL https://onlinelibrary.wiley.com/doi/10.1002/jum.16383.

[15] Umair Khan, Sajjad Afrakhteh, Federico Mento, Noreen Fatima, Laura De Rosa, Leonardo Lucio Custode, Zihadul Azam, Elena Torri, Gino Soldati, Francesco Tursi, Veronica Narvena Macioce, Andrea Smargiassi, Riccardo Inchingolo, Tiziano Perrone, Giovanni Iacca, Libertario Demi, Benchmark methodological approach for the application of artificial intelligence to lung ultrasound data from COVID-19 patients: From frame to prognostic-level, Ultrasonics (ISSN: 0041-624X) 132 (2023) 106994, http://dx.doi.org/10.1016/j.ultras.2023.106994, URL https://www.sciencedirect.com/science/article/pii/S0041624X23000707.

[16] Jiangang Chen, Chao He, Jintao Yin, Jiawei Li, Xiaoqian Duan, Yucheng Cao, Li Sun, Menghan Hu, Wenfang Li, Qingli Li, Quantitative analysis and automated lung ultrasound scoring for evaluating COVID-19 pneumonia with neural networks, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 68 (7) (2021) 2507–2515, http://dx.doi.org/10.1109/TUFFC.2021.3070696, URL https://ieeexplore.ieee.org/document/9393931.

[17] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro Sentelli, Emanuele Peschiera, Riccardo Trevisan, Giovanni Maschietto, Elena Torri, Riccardo Inchingolo, Andrea Smargiassi, Gino Soldati, Paolo Rota, Andrea Passerini, Libertario Demi, Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound, IEEE Trans. Med. Imaging PP (2020) http://dx.doi.org/10.1109/TMI.2020.2994459, URL https://ieeexplore.ieee.org/document/9093068.

[18] Leonardo Lucio Custode, Federico Mento, Francesco Tursi, Andrea Smargiassi, Riccardo Inchingolo, Tiziano Perrone, Libertario Demi, Giovanni Iacca, Multi-objective automatic analysis of lung ultrasound data from COVID-19 patients by means of deep learning and decision trees, Appl. Soft Comput. (ISSN: 1568-4946) 133 (2023) 109926, http://dx.doi.org/10.1016/j.asoc.2022.109926, URL https://www.sciencedirect.com/science/article/pii/S1568494622009759.

[19] Oz Frank, Nir Schipper, Mordehay Vaturi, Gino Soldati, Andrea Smargiassi, Riccardo Inchingolo, Elena Torri, Tiziano Perrone, Federico Mento, Libertario Demi, Meirav Galun, Yonina C. Eldar, Shai Bagon, Integrating domain knowledge into deep networks for lung ultrasound with applications to COVID-19, IEEE Trans. Med. Imaging 41 (3) (2022) 571–581, http://dx.doi.org/10.1109/TMI.2021.3117246, URL https://ieeexplore.ieee.org/document/9557273.

[20] Roshan Roshankhah, Yasamin Karbalaeisadegh, Hastings Greer, Federico Mento, Gino Soldati, Andrea Smargiassi, Riccardo Inchingolo, Elena Torri, Tiziano Perrone, Stephen Aylward, et al., Investigating training-test data splitting strategies for automated segmentation and scoring of COVID-19 lung ultrasound images, J. Acoust. Soc. Am. 150 (6) (2021) 4118–4127, URL https://pubs.aip.org/asa/jasa/article/150/6/4118/993699/Investigating-training-test-data-splitting.

[21] Umair Khan, Sajjad Afrakhteh, Federico Mento, Gizem Mert, Andrea Smargiassi, Riccardo Inchingolo, Francesco Tursi, Veronica Narvena Macioce, Tiziano Perrone, Giovanni Iacca, et al., Low-complexity lung ultrasound video scoring by means of intensity projection-based video compression, Comput. Biol. Med. 169 (2024) 107885, URL https://www.sciencedirect.com/science/article/pii/S0010482523013501.

[22] Umair Khan, Federico Mento, Lucrezia Nicolussi Giacomaz, Riccardo Trevisan, Andrea Smargiassi, Riccardo Inchingolo, Tiziano Perrone, Libertario Demi, Deep learning-based classification of reduced lung ultrasound data from COVID-19 patients, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 69 (5) (2022) 1661–1669, URL https://ieeexplore.ieee.org/document/9740147.

[23] Leonardo Carrer, Elena Donini, Daniele Marinelli, Massimo Zanetti, Federico Mento, Elena Torri, Andrea Smargiassi, Riccardo Inchingolo, Gino Soldati, Libertario Demi, et al., Automatic pleural line extraction and COVID-19 scoring from lung ultrasound data, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 67 (11) (2020) 2207–2217, URL https://ieeexplore.ieee.org/document/9127515.

[24] Markus Lerchbaumer, Jonathan Lauryn, Ulrike Bachmann, Philipp Enghard, Thomas Fischer, Jana Grune, Niklas Hegemann, Dmytro Khadzhynov, Jan Kruse, Lukas Lehner, Tobias Lindner, Timur Özkan, Daniel Zickler, Wolfgang Kuebler, Bernd Hamm, Kai-Uwe Eckardt, Frédéric Muench, Point-of-care lung ultrasound in COVID-19 patients: inter- and intra-observer agreement in a prospective observational study, Sci. Rep. 11 (2021) http://dx.doi.org/10.1038/s41598-021-90153-2, URL https://www.nature.com/articles/s41598-021-90153-2.

[25] Wufeng Xue, Chunyan Cao, Jie Liu, Yilian Duan, Haiyan Cao, Jian Wang, Xumin Tao, Zejian Chen, Meng Wu, Jinxiang Zhang, Hui Sun, Yang Jin, Xin Yang, Ruobing Huang, Feixiang Xiang, Yue Song, Manjie You, Wen Zhang, Lili Jiang, Ziming Zhang, Shuangshuang Kong, Ying Tian, Li Zhang, Dong Ni, Mingxing Xie, Modality alignment contrastive learning for severity assessment of COVID-19 from lung ultrasound and clinical information, Med. Image Anal. (ISSN: 1361-8415) 69 (2021) 101975, http://dx.doi.org/10.1016/j.media.2021.101975, URL https://www.sciencedirect.com/science/article/pii/S1361841521000219.

[26] Federico Mento, Umair Khan, Francesco Faita, Andrea Smargiassi, Riccardo Inchingolo, Tiziano Perrone, Libertario Demi, State of the art in lung ultrasound, shifting from qualitative to quantitative analyses, Ultrasound Med. Biol. (ISSN: 0301-5629) 48 (12) (2022) 2398–2416, http://dx.doi.org/10.1016/j.ultrasmedbio.2022.07.007, URL https://www.sciencedirect.com/science/article/pii/S0301562922004823.

[27] Kaustav Mohanty, John Blackwell, Thomas Egan, Marie Muller, Characterization of the lung parenchyma using ultrasound multiple scattering, Ultrasound Med. Biol. (ISSN: 0301-5629) 43 (5) (2017) 993–1003, http://dx.doi.org/10.1016/j.ultrasmedbio.2017.01.011, URL https://www.sciencedirect.com/science/article/pii/S0301562917300170.

[28] Gino Soldati, Andrea Smargiassi, Libertario Demi, Riccardo Inchingolo, Artifactual lung ultrasonography: It is a matter of traps, order, and disorder, Appl. Sci. (ISSN: 2076-3417) 10 (5) (2020) http://dx.doi.org/10.3390/app10051570, URL https://www.mdpi.com/2076-3417/10/5/1570.

[29] Federico Mento, Gino Soldati, Renato Prediletto, Marcello Demi, Libertario Demi, Quantitative lung ultrasound spectroscopy applied to the diagnosis of pulmonary fibrosis: first clinical study, IEEE Trans. Ultrason. Ferroelectr. Freq. Control PP (2020) 1, http://dx.doi.org/10.1109/TUFFC.2020.3012289.

[30] Mattia Perpenti, Federico Mento, Sajjad Afrakhteh, Giuliana Barcellona, Tiziano Perrone, Libertario Demi, A novel empirical wavelet transform approach for classification of radiofrequency lung ultrasound signals applied to diagnosis of lung diseases, in: 2024 IEEE Symposium (UFFC-JS), IEEE, 2024, pp. 1–4, URL https://ieeexplore.ieee.org/document/10793884.

[31] Federico Mento, Mattia Perpenti, Giuliana Barcellona, Tiziano Perrone, Libertario Demi, Lung ultrasound spectroscopy applied to the differential diagnosis of pulmonary diseases: an in vivo multicenter clinical study, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2024) URL https://doi.org/10.1109/TUFFC.2024.3454956.

[32] Thomas Koudstaal, Marlies S. Wijsenbeek, Idiopathic pulmonary fibrosis, La Press. Médicale (ISSN: 0755-4982) 52 (3) (2023) 104166, http://dx.doi.org/10.1016/j.lpm.2023.104166, URL https://www.sciencedirect.com/science/article/pii/S0755498223000039. Rare respiratory diseases.

[33] Boran Zhou, Brian J. Bartholmai, Sanjay Kalra, Xiaoming Zhang, Predicting lung mass density of patients with interstitial lung disease and healthy subjects using deep neural network and lung ultrasound surface wave elastography, J. Mech. Behav. Biomed. Mater. (ISSN: 1751-6161) 104 (2020) 103682, http://dx.doi.org/10.1016/j.jmbbm.2020.103682, URL https://www.sciencedirect.com/science/article/pii/S1751616119315930.

[34] Toru Kameda, Nao Kamiyama, Nobuyuki Taniguchi, The mechanisms underlying vertical artifacts in lung ultrasound and their proper utilization for the evaluation of cardiogenic pulmonary edema, Diagnostics 12 (2022) http://dx.doi.org/10.3390/diagnostics12020252, URL https://www.mdpi.com/2075-4418/12/2/252.

[35] Federico Mento, Matteo Perini, Ciro Malacarne, Libertario Demi, Ultrasound multifrequency strategy to estimate the lung surface roughness, in silico and in vitro results, Ultrasonics (ISSN: 0041-624X) 135 (2023) 107143, http://dx.doi.org/10.1016/j.ultras.2023.107143, URL https://www.sciencedirect.com/science/article/pii/S0041624X23002196?via%3Dihub.

[36] Mattia Perpenti, Federico Mento, Giovanni Pierro, Alessandro Perrotta, Andrea Smargiassi, Riccardo Inchingolo, Libertario Demi, Novel quantitative lung ultrasound spectroscopy approach for diseases classification, in: 2024 IEEE (UFFC-JS), 2024, pp. 1–4, http://dx.doi.org/10.1109/UFFC-JS60046.2024.10793543, URL https://ieeexplore.ieee.org/document/10793543.

[37] Libertario Demi, Marcello Demi, Renato Prediletto, Gino Soldati, Real-time multi-frequency ultrasound imaging for quantitative lung ultrasound - first clinical results, J. Acoust. Soc. Am. 148 (2020) 998–1006.

[38] Piero Tortoli, Luca Bassi, Enrico Boni, Alessandro Dallai, Francesco Guidi, Stefano Ricci, ULA-OP: An advanced open platform for ultrasound research, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2009) http://dx.doi.org/10.1109/TUFFC.2009.1303.

[39] Federico Mento, Tiziano Perrone, Anna Fiengo, Veronica Macioce, Andrea Smargiassi, Riccardo Inchingolo, Libertario Demi, Limiting the areas inspected by lung ultrasound leads to an underestimation of COVID-19 patients' condition, Intensive Care Med. 47 (2021) http://dx.doi.org/10.1007/s00134-021-06407-0, URL https://link.springer.com/article/10.1007/s00134-021-06407-0.

[40] Libertario Demi, Wim Hoeve, Ruud van Sloun, Gino Soldati, Marcello Demi, Determination of a potential quantitative measure of the state of the lung using lung ultrasound spectroscopy, Sci. Rep. 7 (2017) http://dx.doi.org/10.1038/s41598-017-13078-9, URL https://www.nature.com/articles/s41598-017-13078-9.

[41] Ryan J. Urbanowicz, Melissa Meeker, William La Cava, Randal S. Olson, Jason H. Moore, Relief-based feature selection: Introduction and review, J. Biomed. Informatics (ISSN: 1532-0464) 85 (2018) 189–203, http://dx.doi.org/10.1016/j.jbi.2018.07.014, URL https://www.sciencedirect.com/science/article/pii/S1532046418301400.

[42] Hana Alsomali, Evelyn Palmer, Avinash Aujayeb, Wendy Funston, Early diagnosis and treatment of idiopathic pulmonary fibrosis: A narrative review, Pulm. Ther. 9 (2023) http://dx.doi.org/10.1007/s41030-023-00216-0, URL https://link.springer.com/article/10.1007/s41030-023-00216-0.

[43] Christian Zanza, Francesco Saglietti, Manfredi Tesauro, Yaroslava Longhitano, Gabriele Savioli, Mario Giosuè Balzanelli, Tatsiana Romenskaya, Luigi Cofone, Ivano Pindinello, Giulia Racca, et al., Cardiogenic pulmonary edema in emergency medicine, Adv. Respir. Med. 91 (5) (2023) 445–463, URL https://www.mdpi.com/2543-6031/91/5/34.

[44] Elke Platz, Pardeep S. Jhund, Ross T. Campbell, John J. McMurray, Assessment and prevalence of pulmonary oedema in contemporary acute heart failure trials: a systematic review, Eur. J. Hear. Fail. 17 (9) (2015) 906–916, http://dx.doi.org/10.1002/ejhf.321, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ejhf.321, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.321.

[45] World Health Organization, Pneumonia in children, 2022, URL https://www.who.int/news-room/fact-sheets/detail/pneumonia.

[46] America Lung Association, Pulmonary fibrosis progression and exacerbation, 2024, URL https://www.lung.org/lung-health-diseases/lung-disease-lookup/pulmonary-fibrosis/patients/living-well-with-pulmonary-fibrosis/progression-and-exacerbation.

[47] World Health Organization, Chronic obstructive pulmonary disease (COPD), 2024, URL https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd).

[48] Federico Mento, Mattia Perpenti, Giuliana Barcellona, Tiziano Perrone, Libertario Demi, Quantitative lung ultrasound spectroscopy classification performance in differentiating CPE, pneumonia, and PF, a comparative classifiers' analysis, in: 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS), 2024, pp. 1–4, http://dx.doi.org/10.1109/UFFC-JS60046.2024.10793564.

[49] Gino Soldati, Andrea Smargiassi, Riccardo Inchingolo, Danilo Buonsenso, Tiziano Perrone, Domenica Briganti, Stefano Perlini, Elena Torri, Alberto Mariani, Elisa Mossolani, Francesco Tursi, Federico Mento, Libertario Demi, Proposal for international standardization of the use of lung ultrasound for COVID-19 patients; a simple, quantitative, reproducible method, J. Ultrasound Med. 39 (2020) http://dx.doi.org/10.1002/jum.15285.