

RESEARCH

Open Access



Automatic deep learning-based pleural effusion segmentation in lung ultrasound images

Damjan Vukovic^{1,2*†}, Andrew Wang^{3†}, Maria Antico^{1,4}, Marian Steffens¹, Igor Ruvinov¹, Ruud JG van Sloun⁵, David Canty^{3,6}, Alistair Royse³, Colin Royse^{3,7}, Kavi Haji³, Jason Dowling⁴, Girija Chetty⁸ and Davide Fontanarosa^{1,2*}

Abstract

Background Point-of-care lung ultrasound (LUS) allows real-time patient scanning to help diagnose pleural effusion (PE) and plan further investigation and treatment. LUS typically requires training and experience from the clinician to accurately interpret the images. To address this limitation, we previously demonstrated a deep-learning model capable of detecting the presence of PE on LUS at an accuracy greater than 90%, when compared to an experienced LUS operator.

Methods This follow-up study aimed to develop a deep-learning model to provide segmentations for PE in LUS. Three thousand and forty-one LUS images from twenty-four patients diagnosed with PE were selected for this study. Two LUS experts provided the ground truth for training by reviewing and segmenting the images. The algorithm was then trained using ten-fold cross-validation. Once training was completed, the algorithm segmented a separate subset of patients.

Results Comparing the segmentations, we demonstrated an average Dice Similarity Coefficient (DSC) of 0.70 between the algorithm and experts. In contrast, an average DSC of 0.61 was observed between the experts.

Conclusion In summary, we showed that the trained algorithm achieved a comparable average DSC at PE segmentation. This represents a promising step toward developing a computational tool for accurately augmenting PE diagnosis and treatment.

Keywords Point-of-Care Ultrasound, Lung Ultrasound, Deep Learning, Pleural Effusion / diagnostic imaging

[†]Damjan Vukovic and Andrew Wang contributed equally to this work (shared first authors).

*Correspondence:

Damjan Vukovic
n10116427@qut.edu.au
Davide Fontanarosa
d3.fontanarosa@qut.edu.au

¹ School of Clinical Sciences, Queensland University of Technology, Gardens Point Campus, 2 George St, Brisbane QLD 4000, Australia

² Centre for Biomedical Technologies (CBT), Queensland University of Technology, Brisbane QLD 4000, Australia

³ Department of Surgery (Royal Melbourne Hospital), University of Melbourne, Royal Parade, Parkville VIC 3050, Australia

⁴ CSIRO Health and Biosecurity, The Australian eHealth Research Centre, Herston QLD 4029, Australia

⁵ Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

⁶ Department of Medicine and Nursing, Monash University, Wellington Road, Clayton 3800, Victoria, Australia

⁷ Outcomes Research Consortium, Cleveland Clinic, Cleveland, Ohio, USA

⁸ School of IT & Systems, Faculty of Science and Technology, University of Canberra, 11 Kirinari Street, Bruce ACT 2617, Australia



Introduction

Pleural effusion (PE), the excessive accumulation of fluid in the pleural cavity, is the most common disease among all pleural diseases [1]. The most cited population study (Czech Republic) measured an incidence of 0.32 percent annually for PE [2]. In patients with COVID-19, the overall incidence of PE was 7.3 percent [3]. While history taking, a physical examination remains the basis for evaluating patients with PE, the British Thoracic Society in its latest draft guideline on pleural disease [4] has included a recommendation on performing lung ultrasound (LUS) on every patient's initial presentation and when procedures involving the pleura are being performed. LUS provides a higher accuracy than physical examination [5] and chest radiography in the detection of PE [6]. However, LUS assessment typically involves a highly trained and experienced operator acquiring the images using a well-outlined protocol [7] and interpreting the images for pattern changes associated with lung pathology. The conventional LUS assessment, therefore, requires considerable time and resources [8].

Beyond recognizing and identifying PE, the underlying causes of PE are numerous, and an invasive procedure known as thoracentesis is frequently required to guide differential diagnosis and treatment [4]. The LUS-guided intervention has been shown to increase success at obtaining pleural fluid and, in turn, reduce the risk of complications such as pneumothorax [9]. This procedure is rapidly becoming standard patient care. To perform thoracentesis, recognition of the extent and location of the PE is critical, as a needle or catheter is introduced to the pleural space. A previous study from our group [10] implemented a deep learning (DL) based algorithm to detect the presence of PE on LUS. We demonstrated a comparable accuracy between the algorithm and the LUS experts [10]. This was the first step in supporting clinicians to identify image patterns associated with PE. Recognizing that LUS currently demands significant resources, and that LUS is highly desirable in patients with PE, we propose a DL-based algorithm to automatically segment PE to assist clinician(s) in localizing PE in LUS videos for safe needle guidance during thoracentesis procedures, improve PE localization accuracy, and aid clinicians during PE volume estimation. This automated algorithm for segmenting PE could provide considerable benefits to initial differential diagnosis, subsequent patient safety, and improved procedural success if a more invasive diagnostic and treatment approach is warranted. The performance of the proposed algorithm was evaluated using the Dice Similarity Coefficient (DSC) and the complementary Szymkiewicz-Simpson Overlap Similarity Coefficient (OVC) against the LUS experts. The inter-observer variability between the LUS experts' contours

and the algorithms' performance on PE segmentations were evaluated using the DSC and OVC metrics.

This study's central aim is to confront the current hurdles in developing an automated segmentation method for PE detection, a previously unexplored territory. Additionally, it builds upon prior work [11] by delving into the assessment of segmentation variations among experts. Existing research on LUS PE segmentation is limited, with only one study available, which does not perform segmentation on a per-patient basis where the frames are not divided into the training, validation and test set per patient. Additionally, this study relies solely on average Dice scores for the entire dataset and lacks an evaluation metric that assesses the common overlapped area between segmentations, such as the Szymkiewicz-Simpson overlap coefficient.

This study aims to address these gaps by establishing a comprehensive LUS scanning protocol that covers various potential lung pathologies, with a particular focus on addressing PE as an initial step. The deep learning approach builds on previous classification work and addresses the need for automated segmentation of imaging patterns and artifacts in the lung that indicate PE. This segmentation enables the visual presentation of PE regions to sonographers, facilitating the diagnostic process of fluid drainage in the lung (thoracentesis). Furthermore, it lays the foundation for potential sonographer image guidance in future applications and research.

Related works

The ability to segment LUS images has been explored for several respiratory pathologies. Most notably, Morilhat et al. [12] recently shared their findings from a DL-based algorithm (nnU-net) for PE segmentation with median DSC scores of 0.74 and 0.82 for two datasets consisting of patients from low-to-middle-income countries suspected of tuberculosis. The group performed an automatic DL-based segmentation using spatial information consisting of 2D pixel coordinate information as an additional input into their algorithm. Their algorithm performance was evaluated on a per-dataset basis and compared to the experts' through an interobserver study. Our study differs from Morilhat et al. [12] in several aspects, including a different patient cohort based on an ultrasound (US) imaging acquisition protocol that focused on LUS pathology identification, DSC/OVC evaluations done on a per patient basis using an improved clinical contouring criteria, US probe settings, and a CNN (with Reg-STN) based algorithm.

The other studied LUS image patterns included COVID-19 markers [13–15] for producing an automated COVID-19 associated pathologies segmentation tool. Roy et al. [14] implemented a weakly supervised

classification algorithm using a Spatial Transform Network (STN) and a segmentation algorithm to identify and spatially localize pathology image patterns associated with COVID-19 markers in LUS videos, respectively. In addition, Mento et al. [13] used the same algorithm on new sets of COVID-19 data and applied a threshold technique to the experts' and algorithm's frame segmentations to improve the video-based agreement between the algorithm predictions and the experts' ground truths. In contrast, Roshan et al. [15] performed automatic segmentation of COVID-19 associated imaging patterns using a modified U-net structure and included the OVC to further determine the agreement between the predicted and the ground truth segmentations.

Materials and methods

This study was approved by The Melbourne Health Human Research Ethics Committee (Australia) (28/08/2018, ACTRN12618001442291, HREC/66935/MH-2020) and was performed in accordance with the Declaration of Helsinki. LUS images used in this study were acquired from a previous study [16]. Written informed consent was obtained from all participating patients [16] (Melbourne Health Human Research Ethics Committee approval HREC/18/MH/269, trial registration: <http://www.ANZCTR.org.au/ACTRN12618001442291.aspx>).

All patients were admitted to the Royal Melbourne Hospital under an Internal Medicine unit with a cardiorespiratory-related presentation. The LUS examination was performed by an experienced physician trained (XC) in point-of-care US (POCUS) [16] using a Sonosite X-Porte portable US imaging system (Fujifilm, Bothell, WA, USA) with settings shown in Table 1 [17]. The examination followed a standardized iLungScan protocol (The University of Melbourne, Ultrasound Education Group [17]). Patients were positioned in a supine position, and six distinct lung scanning zones [18] were examined at least once (Fig. 1). Images were immediately reviewed for diagnostic accuracy and quality assurance by a second LUS expert (DC, AR, or CR).

A total of 51 LUS videos from 24 patients were used to train our automated PE segmentation algorithm. The algorithm was trained on approximately an equal number of expert 1 and expert 2 ground truths during hyperparameter optimization. The cross-validation was performed on a per patient level using each expert's segmentation labels and calculating the DSC/OVC

Table 1 Ultrasound imaging system parameters [17]

Sonosite X-Porte	
Probe/Transducer	Phased Array
Preset	Cardiac
Frequency	1-5 MHz
Acoustic working Frequency	1.72 MHz
Mechanical index	1.3
Soft-tissue thermal index	0.9
Focal optimization	Gen
Default penetration depth	15 cm
Pulse Repetition	2933 Hz
Scan repetition rate	34.1 Hz
Tissue harmonic imaging	On

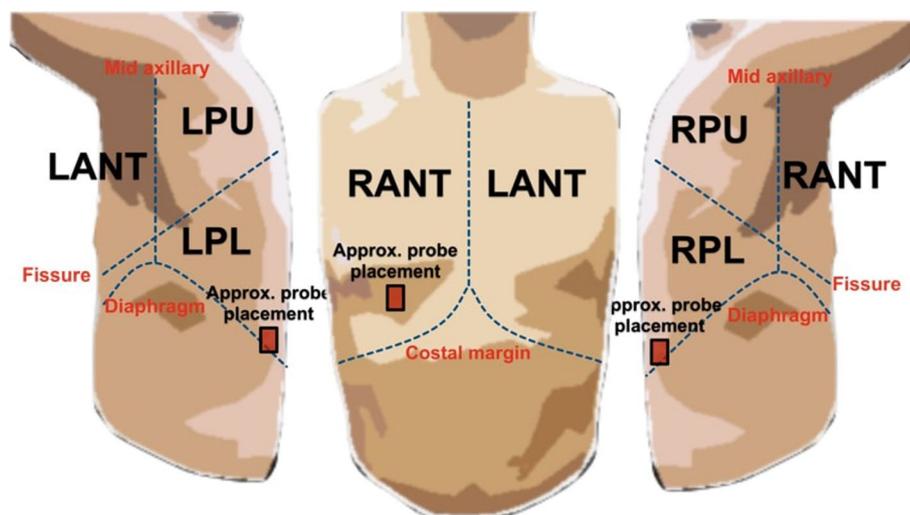


Fig. 1 Examples of the scanning regions (viz. Right Anterior (RANT), Right Posterior Lower (RPL), Left Posterior Lower (LPL)) and the approximate probe placement during the image acquisition of LUS frames containing PE. **Figure created and owned by coauthors

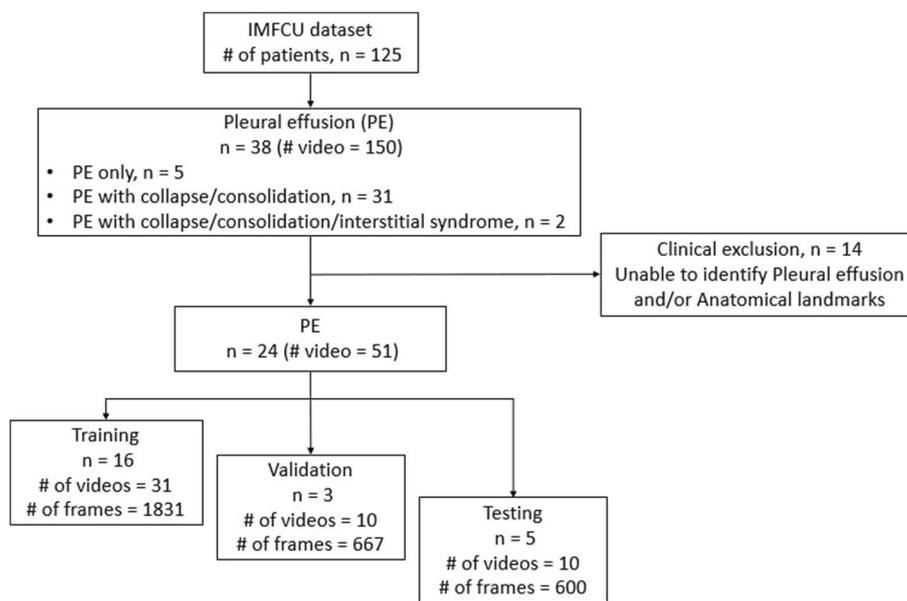


Fig. 2 A flow diagram showing the pathology distribution of the original PE patients to the final training/validation/test dataset split used to train the algorithm

Table 2 Breakdown of PE patients during algorithm training/testing

	# Patients	# Videos	# Frames
Effusion	3	4	240
Collapse / Consolidation / Effusion	20	43	2561
Acute Pulmonary Oedema / Interstitial Syndrome / Effusion	1	4	240
Total	24	51	3041

scores between the experts and the algorithm-predicted segmentation.

Dataset

The 38 PE patient dataset (detailed in Fig. 2) was obtained based on a six-scanning region LUS image acquisition protocol [17, 18] shown in Fig. 1. These six scanning regions include the Left Anterior (LANT), Right Anterior (RANT), Left Posterior Upper (LPU), Left Posterior

Lower (LPL), Right Posterior Upper (RPU), and Right Posterior Lower (RPL) regions.

The dataset from Table 2 has been further analyzed, according to the clinical criteria consisting of the experts’ (trained sonographer (MS) and trained LUS MD (AW)) ability to identify PE imaging patterns, which is further detailed in “Frame-based contouring strategy” section. As a result, the number of PE patients was reduced from 38 to 24 as shown in Fig. 2. Table 2 outlines the number of patients, videos and frames for each of the pathologies that presented alongside PE and is the dataset that is used to train and test our algorithm.

Table 3 shows the scanning regions and the number of videos/frames associated with the patient pathology distribution from Table 2. The PE dataset breakdown from Tables 2 and 3 show that PE patients present with multiple other pathologies (Consolidation/Collapse, APO, Interstitial Syndrome) other than PE. A significant representation of the dataset contains patients with a combined diagnosis of PE/Consolidation/Collapse located in the RPL and LPL LUS scanning regions.

Table 3 The patients from Table 2 and their respective scanning region distributions shown per video and per frame

	RANT	RPL	LPL
Effusion	N/A	# Video(s) 3 # Frame(s) 180	# Video(s) 1 # Frame(s) 60
Collapse / Consolidation / Effusion	# Video(s) 1 # Frame(s) 60	# Video(s) 15 # Frame(s) 900	# Video(s) 27 # Frame(s) 1601
Acute Pulmonary Oedema / Interstitial Syndrome / Effusion	N/A	# Video(s) 1 # Frame(s) 60	# Video(s) 3 # Frame(s) 180

Pre-processing

The anonymization and removal of unique patient identifiers on the LUS videos with imaging patterns associated with PE was done before any segmentation labels could be completed. The next step involved extracting the LUS videos from the compressed Digital Imaging and Communications in Medicine [19] (DICOM) image format using Pydicom [20] and removing overlays (e.g., text, watermarks, trademarks, etc.) outside the US sector. The image pixel dimensions were $(0.02 \cdot 0.02)mm$. The original image size $(960 \cdot 720)pixels$ was reduced to $(806 \cdot 550)pixels$ by cropping the images to enclose the US sector, thus minimizing the presence of black pixels surrounding the relevant image information.

Clinical PE contouring methods

Following image pre-processing, two reviewers (AW and MS) trained in LUS were assigned LUS videos in DICOM format for independent image interpretation. The reviewers did not have knowledge of the extent of PE on a video or frame level before reviewing the videos. The reviewers used a modified version of the Labelme [21] program to view and outline the PE. Sixty consecutive frames were selected from the videos most representative of the pathology at the reviewers' discretion. Polygons were created to outline the PE on a frame-by-frame basis. The coordinates of the polygons were recorded in the open standard file format of JavaScript Object Notation [22] (JSON).

Identification of PE on LUS followed the Lung Ultrasound Interpretation Score protocol developed at the University of Melbourne Ultrasound Education Group, which is based on the international evidence-based LUS recommendations from Volpicelli et al. [23].

Figure 3 provides a visual representation of the exclusion criteria used in this study. In Image A, we can observe an LUS frame that is not suitable for training purposes due to several reasons, primarily centered around poor image quality. These issues include the absence of discernible anatomical or pathological markers. Image A represents one of the extreme cases of an inconclusive frame.

On the other hand, Image B in Fig. 3 illustrates an ideal LUS frame or video for PE identification. Here, the clinical significance lies in the clear visibility and identification of anatomical markers such as the collapsed lung, pleural lining, and diaphragm, along with the presence of PE-associated imaging patterns.

The LUS reviewers also outlined the PE using identifiable anatomical features, including the diaphragm, collapsed or consolidated lung, the chest wall, and the clear line that delineated the extent of the image sector. On

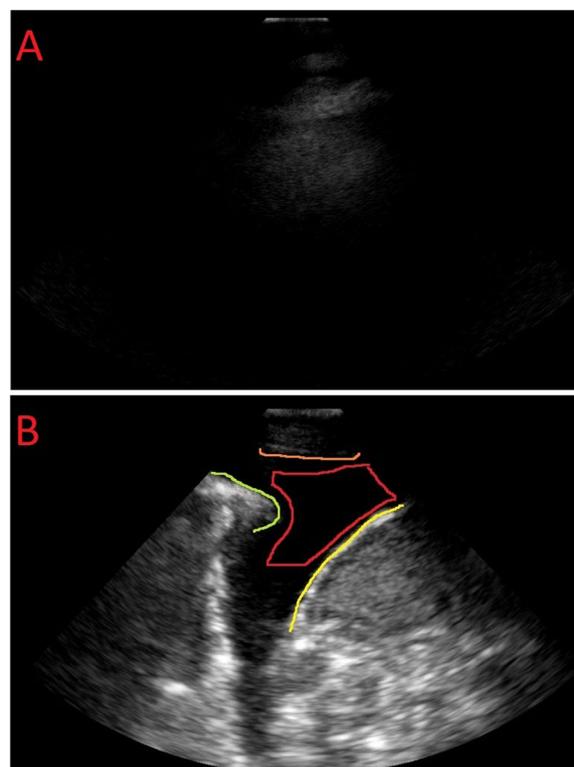


Fig. 3 Comparison of PE identified LUS frames, where image (A) does not meet the clinical criteria needed for contouring PE and image (B) is the ideal case for PE segmentation. The presence of PE was defined as an anechoic space (in red) between the parietal pleura (lining the chest wall in orange and diaphragm in yellow) and visceral pleura (lining the lung surface in green) image B, along with the collapsed lung tissue (in green)

completing one video frame, the reviewers repeated the same process on the subsequent frames. The same polygons could be transferred from the previous frame to the subsequent one, accounting for the pathology being continuous in the consecutive frames. However, the reviewers were required to examine each frame closely to ensure the polygon delineating PE on the previous frame remained applicable to the following frame. Depending on the amount of respirophasic movement of the lung and effusion, the reviewers had to adjust the polygons by applying the same diagnostic principle.

The contours for the final training and validation sets consisted 50% from expert 1 and 50% from expert 2. However, in the testing set, each patient had contours created by both experts.

Frame-based contouring strategy

After application of the clinical data exclusion criteria, each LUS frame containing PE imaging patterns had an associated contour that was used as the ground truth segmentation label during training and assessment of

the algorithm performance. These LUS contoured frames were then split into training, validation and testing sets based on the number of patients, and were divided into their respective videos, frames, and scanning regions as shown in Table 4. These training/testing dataset splits are based on the standard 80/20 split [24] to balance the variance between the training and testing performance.

Deep learning model

The approach used leveraged a DL architecture that combined two key components: a Convolutional Neural Network (CNN) and a Spatial Transformer Network (STN) [14] into our framework to accurately pinpoint the presence of pulmonary pleural effusions (PE). Additionally, the CNN segment of our algorithm was built upon a customized U-net architecture.

This architecture allowed us to effectively identify and localize PE within lung ultrasound scans, contributing to the diagnostic process by automating the detection of this pathology. The STN played a crucial role in precisely mapping the regions of interest, while the CNN provided the necessary segmentation and classification capabilities for accurate detection.

The algorithm was trained by minimizing the pixel-wise categorical cross-entropy loss between the segmentation generated from the clinicians’ contours (ground truths) and the segmentations predicted by the algorithm [14]. Binary semantic segmentation was used, where one class label is used to represent the background of the LUS frame (pixels that have no PE segmentation), and the second class includes the segmentation that contains the image patterns associated with the PE pathology within a LUS frame.

Training approach

The Weights & Biases [25] framework was utilised to perform hyperparameter tuning based on the

validation set performance. The optimization of these hyperparameters was based on the Bayesian method [26], where the training was based on minimizing the training loss while taking into consideration the validation loss, validation DSC, and training DSC curves to prevent overfitting (Table 5).

The validation process plays a critical role in determining the optimal hyperparameters of the algorithm to ensure its best performance based on training from a specified dataset. This process ensures the robustness of these hyperparameters when the algorithm is subsequently tested on an independent set, which, in this study, consisted of 5 patients. Keeping consistency across the folds was crucial as it allowed for comprehensive testing of the algorithm’s performance, including evaluation against the interobserver study dataset and additional patients.

To achieve this, a repeated 5-fold cross-validation was conducted for each expert, resulting in a total of 10 folds. In each fold, one patient and their associated videos were replaced in the original test set with a randomly selected, non-repeating patient, labeled as R from the original training/validation set. Subsequently, the algorithm was retrained using the optimal hyperparameters obtained during the previous validation phase.

Table 5 The hyperparameters used to train the PE segmentation model

Batch size	32
Batch normalization	True
Dropout	0.5
Epochs	50
Learning rate (Adam)	10 ⁻⁵
Loss function	Binary Cross Entropy

Table 4 Distribution of PE patients following an 80/20 training/testing split that is further divided by the number of videos and frames used in the training, validation, and test sets

Set Type	# Patients	# Videos			# Frames		
Training	16	31			1831		
		RANT	RPL	LPL	RANT	RPL	LPL
		1	12	18	60	720	1051
Validation	3	10			610		
		RANT	RPL	LPL	RANT	RPL	LPL
		0	1	9	0	60	550
Testing	5	10			600		
		RANT	RPL	LPL	RANT	RPL	LPL
		0	6	4	0	360	240

Table 6 The generated cross-validation folds or repeated 5-fold cross-validation per expert, includes the patients in the training set (comprising the original Training (T) and Validation (V) sets) and the test sets. The first 5 folds (1-5) use the contours from expert 1, and the next 5 folds (6-10) use the contours from expert 2

Fold	Set type	Test set	Expert
	T: Training V: Validation	Patient #	Ground Truths
1	T, V + 1	(R1,2,3,4,5)	Expert 1
2	T, V + 2	(1,R2,3,4,5)	Expert 1
3	T, V + 3	(1,2,R3,4,5)	Expert 1
4	T, V + 4	(1,2,3,R4,5)	Expert 1
5	T, V + 5	(1,2,3,4,R5)	Expert 1
6	T, V + 1	(R6,2,3,4,5)	Expert 2
7	T, V + 2	(1,R7,3,4,5)	Expert 2
8	T, V + 3	(1,2,R8,4,5)	Expert 2
9	T, V + 4	(1,2,3,R9,5)	Expert 2
10	T, V + 5	(1,2,3,4,R10)	Expert 2

It’s important to note that for each fold, the original test set patient was integrated into the training set, and vice versa. This patient’s data was used alongside the ground-truth contours provided by the respective expert associated with the test set. For example as shown in Table 6, in the first fold, the contours created by expert 1 for patient 1 were utilized, and that patient was replaced with R1. In the sixth fold, the contours produced by expert 2 for patient 1 were considered, and that patient was replaced with R6. This entire process involved conducting a 5-fold repeated cross-validation for each expert, resulting in the evaluation of the algorithm on 5 independent and previously unseen test sets, along with the inclusion of 10 non-repeating random patients from the training/validation set, totaling 15 patients assessed in total.

The network was trained on a single Nvidia Titan RTX GPU with 24 GB of memory installed on a workstation running Linux with 128GB of memory. The GPU workstation used an Intel i9-9820X CPU with 20 cores running at 3.30 GHz (Lambda Labs, San Francisco, CA, USA).

Evaluations

The trained models used a frame-based segmentation labelling approach and produced frame-level predictions, which were evaluated against the frame-based segmentation ground truths provided by the 2 independent experts using the DSC [27] and OVC [28] metrics. The same metrics were used to evaluate the ground-truth contours’ variability between the two experts (AW, MS)

(i.e. interobserver study) using the experts’ contours generated for the 5 patients belonging to the test set.

The DSC score measures the intersection between the two segmentations as the ratio between the number of pixels intersecting the two segmentations (multiplied by 2) and the sum of the total number of pixels included in each segmentation. Thus, this metric ranges from 0 to 1, where ‘1’ represents a perfect match between the two segmentations.

The OVC score measures the overlap between 2 finite sets or 2 segmentations, by showing how much (value between 0 and 1) of the smaller segmentation is enclosed or contained within the larger segmentation. This metric accounts for and is sensitive to the relative location of the 2 finite sets (segmentations) while the DSC is not. A small DSC score can result from the segmentations barely intersecting/overlapping one another or from one segmentation being within the other when the size difference between them is significant. The OVC serves as a supplementary evaluation metric when combined with the DSC score and provides useful information in our study when comparing segmentation masks of differing sizes.

Results

The worst to best performing average DSC scores for the test set are shown per fold (Table 7), per video (Table 8), and per patient (Table 9). Where the training of the algorithm (on an equal number of expert 1 and expert 2 ground truths) has demonstrated performance on par with or higher than the experts’ evaluation (i.e., interobserver study).

Figure 5 shows a representative example where there was a high overlap (OVC) between the segmentations generated, in contrast to a low DSC. The high OVC indicates that both experts agree on a common area

Table 7 The average DSC scores per fold (F) between the algorithm’s predicted segmentations and each expert’s segmentation labels, ordered from worst to best DSC

F	Algorithm / Expert 1	F	Algorithm / Expert 2
9	0.662 +/- 0.15	8	0.656 +/- 0.14
10	0.689 +/- 0.13	10	0.669 +/- 0.14
4	0.690 +/- 0.12	5	0.692 +/- 0.11
6	0.692 +/- 0.16	2	0.693 +/- 0.13
7	0.698 +/- 0.11	3	0.695 +/- 0.10
3	0.707 +/- 0.10	7	0.710 +/- 0.12
2	0.721 +/- 0.11	4	0.712 +/- 0.15
1	0.724 +/- 0.12	6	0.747 +/- 0.10
8	0.725 +/- 0.07	1	0.757 +/- 0.10
5	0.727 +/- 0.12	9	0.768 +/- 0.09

Table 8 The average DSC scores per video of the test patients (ranked worst to best) between the algorithm’s predicted segmentations and each expert ground truth segmentation labels. Each patient (P) in the test set included 2 videos labelled as a and b

P	Expert 1 / Expert 2	P	Algorithm / Expert 1	P	Algorithm / Expert 2
4b	0.312 +/- 0.10	4b	0.564 +/- 0.02	4b	0.569 +/- 0.02
1a	0.449 +/- 0.06	2b	0.604 +/- 0.03	1a	0.573 +/- 0.04
2b	0.550 +/- 0.05	4a	0.630 +/- 0.03	1b	0.574 +/- 0.04
1b	0.563 +/- 0.08	1b	0.632 +/- 0.06	4a	0.601 +/- 0.03
5b	0.593 +/- 0.05	1a	0.649 +/- 0.06	2b	0.684 +/- 0.03
5a	0.665 +/- 0.10	5a	0.719 +/- 0.06	5b	0.776 +/- 0.02
4a	0.671 +/- 0.06	5b	0.721 +/- 0.05	5a	0.820 +/- 0.01
3b	0.743 +/- 0.04	2a	0.830 +/- 0.02	3b	0.830 +/- 0.02
2a	0.762 +/- 0.05	3b	0.860 +/- 0.03	3a	0.853 +/- 0.01
3a	0.772 +/- 0.02	3a	0.898 +/- 0.02	2a	0.864 +/- 0.02

Table 9 Average DSC scores for the 5 patients (P) in the test set (ranked from worst to best) are shown for the interobserver study (expert 1 / expert 2) and the algorithms performance computation (algorithm / expert 1 and algorithm / expert 2) against each expert

P	Expert 1 / Expert 2	P	Algorithm / Expert 1	P	Algorithm / Expert 2
4	0.491 +/- 0.20	4	0.586 +/- 0.04	1	0.573 +/- 0.04
1	0.506 +/- 0.09	1	0.627 +/- 0.06	4	0.581 +/- 0.03
5	0.629 +/- 0.09	2	0.714 +/- 0.12	2	0.765 +/- 0.09
2	0.656 +/- 0.12	5	0.714 +/- 0.05	5	0.783 +/- 0.03
3	0.758 +/- 0.03	3	0.870 +/- 0.03	3	0.834 +/- 0.02

where PE is located, whilst there is a mismatch in the size of PE (and, thus, low DSC). This is further shown in Table 11 where video 4b is closer to the best average OVC score. When the segmentations are closer in size as in Fig. 4 and are contained within one another, the

OVC and DSC calculation are similar (high DSC and OVC).

In Table 10, the OVC scores are shown from worst to best between the algorithm and each expert (algorithm / expert 1 and algorithm / expert 2).

In Table 11, the OVC scores are shown for the interobserver study (expert 1 / expert 2) and the algorithm’s performance against each expert (algorithm / expert 1 and algorithm / expert 2).

The algorithm’s OVC score performance per patient and per video in Tables 11 and 12 vs the interobserver study (expert 1 / expert 2) was on par overall and performed higher when compared to the worst performing interobserver patient study (patient 2).

In Fig. 6, the segmentations overlap even for the worst performing OVC cases. In image A, the low OVC score indicates that the low DSC score is due to a slight misalignment, in contrast to Fig. 5 where the segmentations are enclosed within each other, resulting in a high OVC score. OVC paired with DSC provides more information

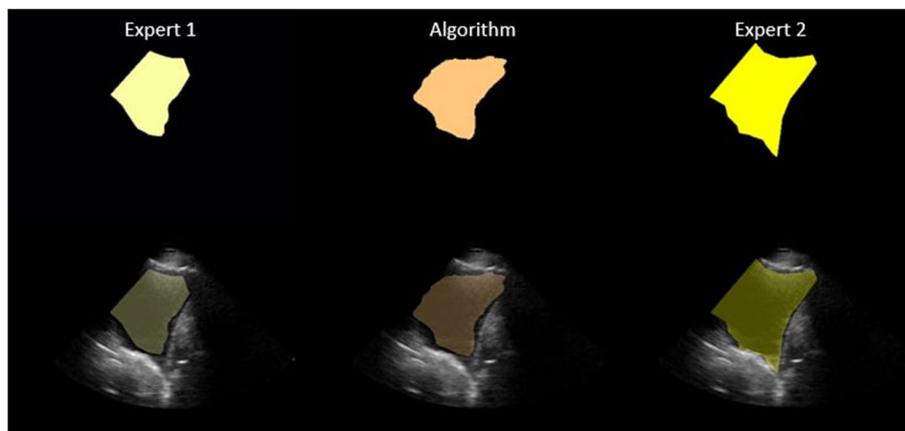


Fig. 4 An example of a LUS image and corresponding segmentations generated by expert 1, expert 2, and the algorithm (from left to right) respectively. On the bottom row, the LUS image and the segmentation are overlaid; on the top row, only segmentations are shown

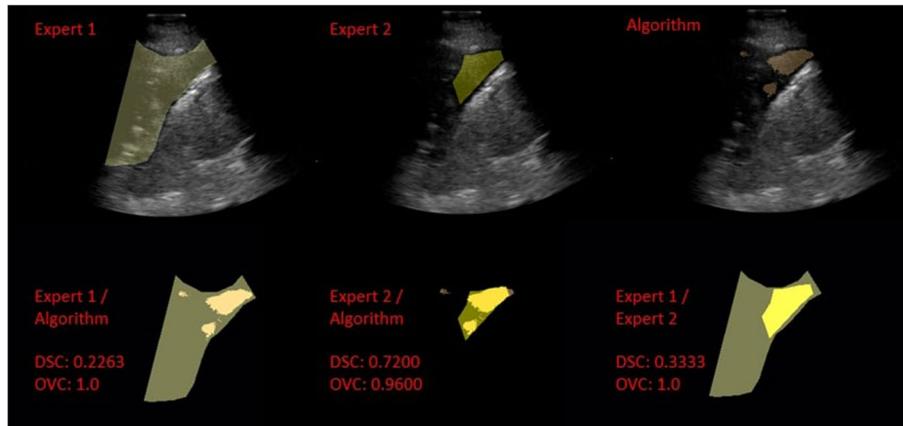


Fig. 5 An example of a LUS image and corresponding segmentations generated by expert 1, expert 2 and the algorithm (from left to right), respectively, extracted from the worst performing video in terms of DSC (i.e., 4b from Table 8). Overlays of LUS image and segmentation are shown in the top row. The bottom row shows the DSC and OVC scores calculated between algorithm / expert 1, algorithm / expert 2, and expert 1 / expert 2 segmentations

Table 10 The OVC scores for the cross-validation results shown per fold (F) from worst to best OVC

F	Algorithm / Expert 1	F	Algorithm / Expert 2
8	0.874 +/- 0.11	4	0.873 +/- 0.10
10	0.877 +/- 0.12	9	0.891 +/- 0.07
5	0.895 +/- 0.12	7	0.902 +/- 0.06
3	0.898 +/- 0.11	10	0.909 +/- 0.10
2	0.908 +/- 0.11	2	0.916 +/- 0.05
6	0.916 +/- 0.12	5	0.920 +/- 0.06
1	0.924 +/- 0.11	1	0.924 +/- 0.05
4	0.932 +/- 0.10	6	0.930 +/- 0.07
7	0.936 +/- 0.09	3	0.931 +/- 0.05
9	0.939 +/- 0.10	8	0.936 +/- 0.03

on the segmentation overlaps, whether it is an enclosure as in Fig. 5 or relatively minor misalignment as in Fig. 6 (images A and B).

Discussion

Our study details a DL algorithm that achieved an average DSC between 0.57 and 0.89 at PE segmentation on LUS images. We demonstrated that its performance was comparable to the experts’ interobserver variability, which ranged between 0.31 and 0.77 DSC.

Interestingly, Morilhat et al. [12] proposed a DL-based PE segmentation model that achieved a median DSC of 0.74 and 0.82 on two datasets acquired using a linear and curvilinear transducer, respectively. Our study differs from Morilhat et al. in several aspects. Firstly, our LUS dataset consisted of images acquired using a phased-array transducer. Secondly, most of our training set consisted of images from the posterior lower zone,

Table 11 The worst to the best average OVC scores shown for the 5 patients (P) and their respective videos (labelled a and b) in the test set

P	Expert 1 / Expert 2	P	Algorithm / Expert 1	P	Algorithm / Expert 2
2b	0.624 +/- 0.10	2b	0.713 +/- 0.02	1a	0.822 +/- 0.09
4a	0.950 +/- 0.09	4a	0.743 +/- 0.02	4a	0.824 +/- 0.04
1b	0.989 +/- 0.03	1b	0.798 +/- 0.06	1b	0.862 +/- 0.09
5a	0.997 +/- 0.01	1a	0.933 +/- 0.04	5a	0.911 +/- 0.02
3a	0.998 +/- 0.01	5b	0.975 +/- 0.01	5b	0.924 +/- 0.04
1a	0.998 +/- 0.01	2a	0.982 +/- 0.01	2b	0.932 +/- 0.02
4b	0.998 +/- 0.01	3b	0.988 +/- 0.01	2a	0.952 +/- 0.03
2a	0.999 +/- 0.01	5a	0.988 +/- 0.01	4b	0.955 +/- 0.01
3b	0.999 +/- 0.01	4b	0.990 +/- 0.01	3b	0.962 +/- 0.02
5b	0.999 +/- 0.01	3a	0.991 +/- 0.01	3a	0.989 +/- 0.01

Table 12 The average OVC scores for the test patients (P) shown from the worst to the best patient

P	Expert 1 / Expert 2	P	Algorithm / Expert 1	P	Algorithm / Expert 2
2	0.812 +/- 0.20	2	0.847 +/- 0.14	1	0.842 +/- 0.09
4	0.974 +/- 0.07	4	0.866 +/- 0.12	4	0.889 +/- 0.07
1	0.994 +/- 0.02	1	0.866 +/- 0.09	5	0.917 +/- 0.03
5	0.998 +/- 0.01	5	0.981 +/- 0.01	2	0.942 +/- 0.03
3	0.999 +/- 0.01	3	0.989 +/- 0.01	3	0.975 +/- 0.02

following the iLungScan protocol [17]. The anatomical landmarks in the posterior lower scanning zone usually include the visualization of the diaphragm. The resulting images may differ from those taken at the BLUE-protocol’s PLAPS-point [29] and the subcostal view. Thirdly, it’s worth noting that most of the pleural effusions in our study exhibited a transudative (anechoic) appearance, primarily because our patients suffered from acute exacerbation of heart failure, and the effusions were located in the lower posterior regions of the lung. This differs from the characteristics of tuberculous pleural effusions, which typically present as exudative (echogenic) in nature. However, it’s important to clarify that the definitive differentiation between exudative and

transudative cases is typically based on clinical criteria and laboratory analysis, such as through needle aspiration, rather than relying solely on ultrasound or other imaging modalities.

Lastly The clinical dataset employed in this study is, admittedly, relatively small in scale. However, despite this constraint, the algorithm showcased its effectiveness in this proof-of-concept study by outperforming the inter-observer study. It effectively identified and localized PE within the LUS images.

In the future, it will be particularly interesting to compare this algorithm and Morilhat et al’s algorithm’s performance on both the tuberculous PE dataset and the internal medicine dataset used in this study.

Regarding other common lung pathologies, a comparison could be made to works from Roy et al. [14]. They recently demonstrated a binary DSC of 0.75 at segmenting COVID-19 associated pathologies, including pleural line abnormalities and consolidations. However, caution should be taken when making the comparison, given the heterogeneity in artifact appearance between PE and the features noted in the COVID-19 study. It is to be noted that our algorithm has been trained to address a single pathology, although concurrent pathologies such as lung collapse/consolidations and interstitial syndrome are common in patients with PE.

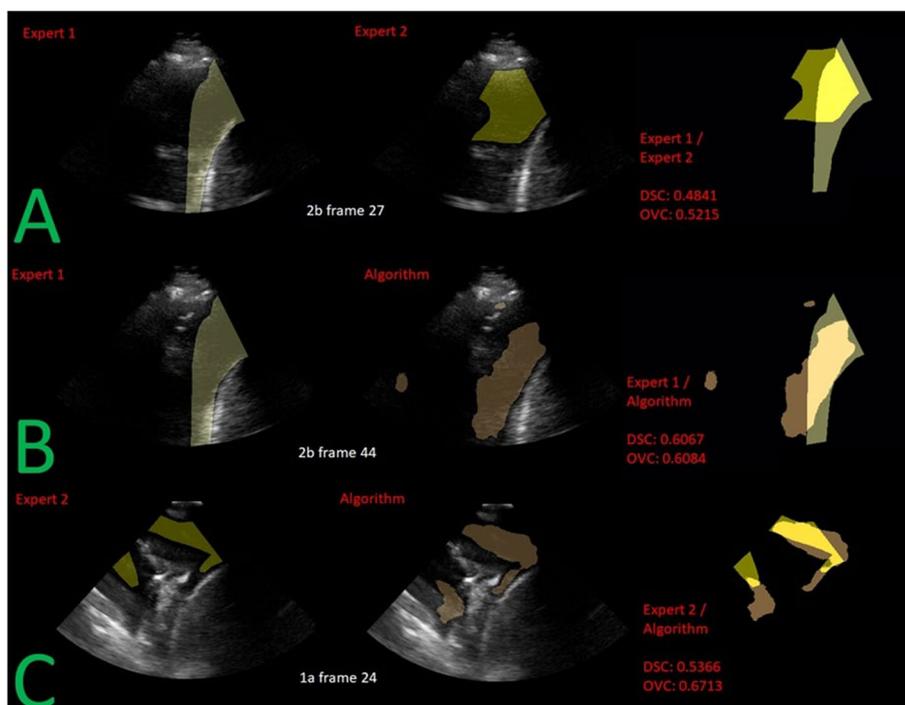


Fig. 6 An example of the corresponding segmentations from expert 1, expert 2 and the algorithm’s prediction shown overlaid over the LUS frame of the worst performing DSC video (i.e., 2b and 1a from Table 11). Each row (A, B, C) shows the worst performing LUS frame of the worst performing video for the interobserver study, algorithm / expert 1, and algorithm / expert 2 respectively

For interobserver variability between the two LUS experts, we reported the lowest average DSC of 0.31 in video 4b (Fig. 5). Video 4b was obtained from the posterior lower zone. To the right of the image were the diaphragm and liver. To the left and in the lower edges of the image, there was echogenicity suggesting the presence of either a collapsed lung, visceral pleura, or complex effusion. Our investigation into the discrepancies of the contours generated by the two experts determined that Expert 1's ground truths likely accounted for images from the preceding and subsequent frames to help determine the extent of the effusion, respectively. In contrast, Expert 2's ground truths were probably restricted to an anechoic region of the highest confidence. In retrospect, the LUS experts likely overestimated and underestimated the extent of the effusion. As expected, since the DL algorithm was trained using the annotations of both experts, it was able to produce a segmentation that was a 'compromise' between the two experts. In fact, the automated segmentation primarily included regions where the two experts agreed (OVC > 0.95 for the worst DSC video). Most notably, the segmentation in video 4b excluded less-certain areas (where the boundaries of PE were unclear) that may be unsafe for the introduction of a needle or catheter during procedures such as ultrasound-guided thoracentesis. Given the significant consequences of injuring the lung tissue or diaphragm, we believe our algorithm is also a step in the right direction for safeguarding patients with PE against iatrogenic complications. The OVC alongside the DSC can be used to determine safe catheter placement during thoracentesis needle insertion.

The highest interobserver DSC reported was 0.77 (e.g. Fig. 4), similar to that observed by Morilhat et al. [12]. Other common measures of interobserver variability are Cohen's kappa and Fleiss' kappa. In a recent COVID-19 lung study, Kumar et al. [30] reported a moderate agreement in the presence (kappa = 0.49) and size of PE (kappa = 0.47). A similar finding was also reported in the paediatric population [31] (kappa = 0.44). It is evident that large variation exists in determining the presence of PE, and agreement remains challenging.

This study has several limitations. First, the study was conducted using LUS images from patients with cardiopulmonary complaints in a single tertiary center. The images were obtained by a single operator using the phased-array transducer from a single manufacturer. The generalizability of the DL algorithm to other clinical and image acquisition settings is to be determined. Second, most of the videos were from the posterior lower scanning zones. A larger sample size that includes effusions detected in various scanning zones, such as the posterior upper zone in a supine patient or anterior zone in a prone

patient, would be beneficial. However, it's worth noting that these factors don't inherently limit the algorithm's performance, as it can identify and localize imaging patterns associated with PE. The variation in imaging/scanning protocols primarily influences the location of PE within the scan, rather than the algorithm's capability to identify and localize these patterns. Where PE incidents are primarily prevalent in unilateral cases [32, 33]. However, they can also occur bilaterally, with most cases being right-sided PE [34]. Notably, right-sided PE cases tend to feature larger effusions compared to the left side [35]. The most common cases of PE are attributed to heart failure [33]. Previous research findings indicate that the location of PE is not an unusual finding, particularly in cases related to heart failure [36]. This study concentrated on an imaging protocol [17] where most cases of PE were located in the lower lung regions, and our algorithm was trained specifically for these scenarios. In the future, we intend to broaden our dataset to ensure the algorithm's effectiveness in identifying this pathology in various lung regions.

Third, the LUS images were reviewed by two LUS experts retrospectively without complete knowledge of the clinical context. This differs from a typical POCUS routine, where the operator acquires and interprets the images to address a specific clinical question. The agreement between experts may improve if the experts acquire and interpret the images in real-time.

The DSC score metric is the standard evaluation metric for segmentation tasks in machine learning (ML). The challenge with the DSC score arises from the need for the size of ground truth and algorithm-predicted segmentations to be the same or relatively similar in size. PE size in LUS images has shown from our study and others [30] that the experts often disagree on the exact size and boundary of PE effusion and that the DSC, at least for PE segmentation, is not sufficient on its own.

Future work may consider the clinical utility of the algorithm output against patient safety. For example, how often the algorithm-outlined PE segmentations are safe for thoracentesis needle insertion. This trained algorithm will also benefit from being tested in other patient cohorts where PE may appear differently (for instance, complex PE with fibrotic materials within the pleural space) to increase the dataset size that could be used for training. Ultimately, a comprehensive tool for PE will likely require consideration of the pretest probability of the pleural effusion and assess the clinical significance of the effusion against the patient's demographic information and clinical history.

Moreover, considering the enhancement of Pleural Effusion (PE) segmentation for improved volume estimation during thoracentesis, the natural progression

involves incorporating spatial information from volumetric Lung Ultrasound (LUS) images into the estimation process. To advance the automatic PE segmentation algorithm, exploration can extend to its application to three-dimensional LUS images. Additionally, various Machine Learning (ML) approaches such as transformers, transfer learning, reinforcement learning, and unsupervised learning can be harnessed to enhance LUS segmentation accuracy in future iterations.

As a preliminary step or alternative approach, consideration can be given to a 2D image fusion technique inspired by Ziyang Zhang et al.'s work [37]. This technique involves using Gaussian pyramids to seamlessly combine 2D data, presenting an intriguing avenue for exploration. Applying this method to the existing dataset could lead to the transformation of a 2D LUS video into a single, larger 2D image. This consolidated image offers a comprehensive view of the scanning region, effectively showcasing the contribution of each frame. Such an approach could prove invaluable for conducting a more detailed and comprehensive analysis of the lung's surface during LUS video assessments.

Conclusion

We proposed an automatic PE segmentation of LUS videos using a DL approach. When compared to experts, we demonstrated the algorithm's capability of segmenting PE on LUS with an average DSC between 0.57 and 0.89. We showed a DSC between 0.31 and 0.77 between the two experts, suggesting a significant degree of variability in PE segmentation. We observed that the algorithm avoided segmenting high-risk, high-uncertainty regions such as potential lung tissues on LUS. These early results are promising for the growing field of ML assisted medicine.

Future focus will be on the generalizability of the algorithm in other settings and datasets, addition of LUS experts in the interobserver study, inclusion of an independent expert(s) to review and score algorithms segmentation based on safety of 'needle placement' during PE drainage procedures (thoracentesis), testing of other DL approaches (transformers, transfer learning, transformers, etc) for PE segmentation, and consideration of volumetric estimation using automated segmentation on three-dimensional ultrasound images.

Acknowledgements

Not applicable.

Authors' contributions

DV and AW- shared coauthors, wrote main manuscript text, performed trainings and clinical image validation. MA, DF - algorithm/clinical feedback. MS - clinical image validation. IR, RS - algorithm feedback, additional trainings. AR, CR, KH, JD, GC - clinical image validation/feedback, clinical discussions. All authors performed revisions/feedback and reviewed the manuscript.

Funding

No funds, grants, or other support were received during the preparation of this manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author(s) on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by The Melbourne Health Human Research Ethics Committee (Australia) (28/08/2018, ACTRN12618001442291, HREC/66935/MH-2020) and was performed in accordance with the Declaration of Helsinki. LUS images used in this study were acquired from a previous study [16].

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 17 January 2023 Accepted: 3 November 2023

Published online: 29 November 2023

References

- Krishna R, et al. Pleural Effusion. In: StatPearls [Internet]. Treasure Island: StatPearls Publishing; 2022. <https://www.ncbi.nlm.nih.gov/books/NBK448189/>. Accessed 5 Jan 2023.
- Marel M, et al. The incidence of pleural effusion in a well-defined region. Epidemiologic study in central Bohemia Chest. 1993. <https://doi.org/10.1378/chest.104.5.1486>.
- Chong WH, et al. The incidence of pleural effusion in COVID-19 pneumonia: state-of-the-art review. Heart Lung. 2021. <https://doi.org/10.1016/j.hrtlng.2021.02.015>.
- Maskell N, et al. BTS guidelines for the investigation of a unilateral pleural effusion in adults. Thorax. 2003. https://doi.org/10.1136/thorax.58.suppl_2.ii8.
- Walsh MH, et al. Comparing accuracy of bedside ultrasound examination with physical examination for detection of pleural effusion. Ultrasound J. 2021. <https://doi.org/10.1186/s13089-021-00241-7>.
- Hansell L, et al. Lung ultrasound has greater accuracy than conventional respiratory assessment tools for the diagnosis of pleural effusion, lung consolidation and collapse: a systematic review. J Physiother. 2021. <https://doi.org/10.1016/j.jphys.2020.12.002>.
- Manivel V, et al. CLUE: COVID-19 lung ultrasound in emergency department. Emerg Med Aust. 2020. <https://doi.org/10.1111/1742-6723.13546>.
- Pietersen I, et al. Lung ultrasound training: a systematic review of published literature in clinical lung ultrasound training. Crit Ultrasound J. 2018. <https://doi.org/10.1186/s13089-018-0103-6>.
- Perazzo A, et al. Can ultrasound guidance reduce the risk of pneumothorax following thoracentesis? J Bras Pneumol. 2014. <https://doi.org/10.1590/S1806-37132014000100002>.
- Tsai CH, et al. Automatic deep learning-based pleural effusion classification in lung ultrasound images for respiratory pathology diagnosis. Physica Medica. 2021;83(38-45). <https://doi.org/10.1016/j.ejmp.2021.02.023>.
- Antico M, et al. Bayesian CNN for Segmentation Uncertainty Inference on 4D Ultrasound Images of the Femoral Cartilage for Guidance in Robotic Knee Arthroscopy. IEEE Access. 2020. <https://doi.org/10.1109/ACCESS.2020.3044355>.
- Morilhat G, et al. Deep Learning-Based Segmentation of Pleural Effusion from Ultrasound Using Coordinate Convolutions. In: Distributed, Collaborative, and Federated Learning, and Affordable AI and Healthcare for Resource Diverse Global Health DeCaF FAIR 2022 2022 Lecture Notes in Computer Science. 2022. https://doi.org/10.1007/978-3-031-18523-6_16.

13. Mento F, et al. Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study. *J Acoust Soc Am*. 2021. <https://doi.org/10.1121/10.0004855>.
14. Roy S, et al. Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound. *IEEE Trans Med Imaging*. 2020. <https://doi.org/10.1109/TMI.2020.2994459>.
15. Roshankhah R, et al. Investigating training-test data splitting strategies for automated segmentation and scoring of COVID-19 lung ultrasound images. *J Acoust Soc Am*. 2021. <https://doi.org/10.1121/10.0007272>.
16. Cid-Serra X, et al. Effect of a Multiorgan Focused Clinical Ultrasonography on Length of Stay in Patients Admitted With a Cardiopulmonary Diagnosis: A Randomized Clinical Trial. *JAMA Netw Open*. 2021. <https://doi.org/10.1001/jamanetworkopen.2021.38228>.
17. Ford JW, et al. A Pilot Assessment of 3 Point-of-Care Strategies for Diagnosis of Perioperative Lung Pathology. *Anesth Analg*. 2017. <https://doi.org/10.1213/ANE.0000000000001726>.
18. Roysse AG, et al. Three Zone Scanning Protocol For Lung Ultrasound: An Anatomical Basis. *Heart Lung Circ*. 2022. <https://doi.org/10.1016/j.hlc.2022.09.013>.
19. Bidgood WD Jr, et al. Introduction to the ACR-NEMA DICOM standard. *Radiographics*. 1992. <https://doi.org/10.1148/radiographics.12.2.1561424>.
20. Mason D, et al. SU-E-T-33: Pydicom: An Open Source DICOM Library. *Med Phys*. 2011. <https://doi.org/10.1118/1.3611983>.
21. Russell BC, et al. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int J Comput Vis*. 2008. <https://doi.org/10.1007/s11263-007-0090-8>.
22. ECMA-404: the JSON data interchange format. ECMA International; 2013. <https://www.iso.org/standard/71616.html>. Accessed 10 Jan 2023.
23. Volpicelli G, et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med*. 2012. <https://doi.org/10.1007/s00134-012-2513-4>.
24. George EP, et al. An Analysis for Unreplicated Fractional Factorials. *Technometrics*. 1986. <https://doi.org/10.1080/00401706.1986.10488093>.
25. Biewald L, et al. Experiment tracking with weights and biases. 2020. <https://www.wandb.com/>. Accessed 10 Jan 2023.
26. Mockus JB, et al. Bayesian approach to global optimization and application to multiobjective and constrained problems. *J Optim Theory Appl*. 1991. <https://doi.org/10.1007/BF00940509>.
27. Dice Lee R, et al. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945. <https://doi.org/10.2307/1932409>.
28. Vijaymeena MK, et al. Clustering with Semantic Similarity for Text Mining. *Middle East J Sci Res*. 2016;24(30–36). <https://doi.org/10.5829/idosi.mejrs.2016.24.TAET23316>.
29. Lichtenstein DA, et al. The BLUE-points: three standardized points used in the BLUE-protocol for ultrasound assessment of the lung in acute respiratory failure. *Crit Ultrasound J*. 2011. <https://doi.org/10.1007/s13089-011-0066-3>.
30. Kumar A, et al. Interobserver Agreement of Lung Ultrasound Findings of COVID-19. *J Ultrasound Med*. 2021. <https://doi.org/10.1002/jum.15620>.
31. Gravel CA, et al. Interrater reliability of pediatric point-of-care lung ultrasound findings. *Am J Emerg Med*. 2020. <https://doi.org/10.1016/j.ajem.2019.01.047>.
32. Antonangelo L, et al. Tuberculous pleural effusion: diagnosis & management. *Expert Rev Respir Med*. 2019. <https://doi.org/10.1080/17476348.2019.1637737>.
33. Bhatnagar R, Nick M, et al. The modern diagnosis and management of pleural effusions. *BMJ*. 2015;351. <https://doi.org/10.1136/bmj.h4520>.
34. Chandela A, et al. Incidence and prognostic significance of pleural effusions in pulmonary arterial hypertension. *Pulm Circ*. 2021. <https://doi.org/10.1177/20458940211012366>.
35. Karkhanis VS, et al. Pleural effusion: diagnosis, treatment, and management. *Emerg Med*. 2012. <https://doi.org/10.2147/OAEM.S29942>.
36. Woodring JH, et al. Distribution of pleural effusion in congestive heart failure: what is atypical? *South Med J*. 2005. <https://doi.org/10.1097/01.SMJ.0000152761.02928.88>.
37. Zhang A, et al. Endoscope image mosaic based on pyramid ORB. *Biomed Signal Process Control*. 2022. <https://doi.org/10.1016/j.bspc.2021.103261>.

Damjan Vukovic Received a B.Sc. in Medical Physics from the University of Windsor, and a M.A.Sc. in Medical Physics from the Queensland University of Technology in 2016 and 2018 respectively. He is currently pursuing a PhD at the Queensland University of Technology where he works on deep learning based approaches to identification and localization of respiratory pathologies using ultrasound imaging.

Andrew Wang Received a M.D in 2017 and a G.C in clinical ultrasound and diagnostic medical sonography/sonographer in 2018 from the University of Melbourne. He is currently pursuing a PhD at the University of Melbourne with interests in Point-of-Care ultrasound, anaesthesia, and peri-operative Medicine.

Maria Antico A research fellow at the CSIRO Australian e-Health Research Centre and an adjunct lecturer at the Queensland University of Technology (Brisbane, Australia). Her research focuses on the automatic interpretation of medical imaging through advanced techniques, such as deep learning techniques, and on the automation of ultrasound scans for medical diagnostics and interventions. Maria has completed her PhD project in 2021 at the Queensland University of Technology. Her PhD project focused on automatic image analysis to provide surgical guidance to autonomous robots in the framework of minimally invasive surgery. Prior to her PhD candidature, she received her BEng. in Engineering Sciences from the University of Rome Tor Vergata (Italy) in 2014 and MEng. in Biomechanical Engineering from the Technical University of Delft (The Netherlands) in 2016, in collaboration with Philips Research (The Netherlands).

Marian Steffens Qualified with a BAppSc - Medical Imaging in 1997 (Queensland University of Technology - QUT, Brisbane, Australia) and worked as a Radiographer at the Royal Brisbane and Women's Hospital whilst completing a Grad Dip - Medical Ultrasound in 2001 (QUT, Brisbane, Australia). She has worked as a Sonographer/Radiographer in several major tertiary referral hospitals in both Australia and UK, including serving over 15 years as a Senior Sonographer/Radiographer at The Prince Charles Hospital, Cherside, Australia - a leading cardio-thoracic hospital. Having trained and collaborated with numerous Medical Imaging Students, Sonographers, Nurses, Medical Consultants & Registrars, she has published and presented for medical imaging professional bodies Australasian Sonographers Association (ASA) and Australian Society of Medical Imaging & Radiation Therapy (ASMIRT). Her current research assistant role at QUT now directs her interest towards reducing user dependence in Ultrasound by using AI integration.

Igor Ruvinov A MASC graduate from the department of Mechanical and Mechatronics Engineering at the University of Waterloo. Prior to that he earned his BASC degree in Nanotechnology Engineering at the University of Waterloo. His research interests include machine learning and its application in various fields. He is currently working as a software engineer.

Ruud JG van Sloun (Member, IEEE) Received the B.Sc. and M.Sc. degrees (cum laude) in electrical engineering and the Ph.D. degree (cum laude) from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2012, 2014, and 2018, respectively. Since then, he has been an Assistant Professor with the Department of Electrical Engineering at the Eindhoven University of Technology and since January 2020 a Kickstart-AI fellow at Philips Research, Eindhoven. From 2019–2020 he was also a Visiting Professor with the Department of Mathematics and Computer Science at the Weizmann Institute of Science, Rehovot, Israel. He is an NWO Rubicon laureate, and received an ERC starting grant and Google Faculty Research Award. His current research interests include deep learning for signal processing and imaging, active signal acquisition, model-based deep

Publisher's Note

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

learning, compressed sensing, ultrasound imaging, and probabilistic signal and image reconstruction.

David Canty Qualified as a cardiothoracic anaesthetist in 2005 (Papworth Hospital, Cambridge, U.K) after completing his medical training in Hobart in 1996 (Tasmania) and anaesthetic training in Melbourne in 2004 (Australia). After commencing as a lecturer and researcher in ultrasound at Ultrasound Education Group, University of Melbourne in 2010 he completed a PhD in the utility of point-of-care transthoracic echocardiography in anaesthesia for noncardiac surgery. His research interests include the utility of point-of-care ultrasound in medicine and surgery, total arterial coronary revascularization and the development of remote and scalable training in point-of-care-ultrasound to meet increasing demand. He has published over 60 publications, three book chapters, achieved over \$4.5M in grants for point-of-care ultrasound and supervises PhD and honours students. He developed a novel scalable and remote method of point-of-care-ultrasound teaching using online learning and ultrasound simulators that have led to the development of a large ultrasound simulation training centre and an increasing number of satellite centres in Australia and North America. More recently, he has applied this method of ultrasound training to the curriculum for medical and physiotherapy students. He was promoted to Associate Professor at the University of Melbourne in 2022. Associate Professor Canty has held the position of chair of the Australian and New Zealand College of Anaesthetists cardiac anaesthesia special interest group for two terms, convening three National scientific meetings. Associate Professor Canty is in his second term as a member of the Society of Cardiovascular Anesthesiologists International Academy and through this association was recently successful in a bid to host the International Congress in Cardiothoracic and Vascular Anesthesia for the first time in Australia (Sydney) in 2025.

Alistair Royse Is a cardiothoracic surgeon and has pioneered the use of arterial coronary bypass grafts and complex arterial reconstructions. This has radically changed the practice of coronary surgery in this country. He has also been intimately involved in the development of clinical ultrasound including intraoperative ultrasound. With his brother, who is an anaesthetist as co-director of the ultrasound education group, they oversee a large distance education (eLearning) program with almost 40,000 students in 3 years. He was promoted via the academic stream to Professor in 2011. He has more than 240 peer reviewed papers.

Colin Royse Qualified as a medical practitioner (M.B.B.S) in 1987 from the University of Melbourne. He subsequently trained as an anaesthesiologist and completed a doctoral research degree (M.D research) in 2000 from the University of Melbourne. He has worked for the University of Melbourne as a teaching and research academic since 2006 and works as a cardiothoracic anesthesiologist at the Royal Melbourne Hospital, Melbourne, Australia. He has continued as an active teaching and research academic focusing on clinical ultrasound, perioperative outcomes, and quality of recovery after surgery. Prof. Royse has authored more than 200 peer-reviewed publications and is co-director of the Ultrasound Education Group at the University of Melbourne, a leading provider of ultrasound education to clinicians.

Kavi Haji Kavi is currently the Intensive Care Unit Clinical Director at Peninsula Health, Melbourne. Kavi has a special interest in ultrasound and echocardiography in critical care and in teaching. She completed a PhD at the University of Melbourne on the role of ultrasound and echocardiography in critical care in 2018. In 2020, Kavi was appointed to the adjunct associate professor position at the University of Melbourne. Kavi is an instructor for the Melbourne

University Ultrasound Education Group. Kavi is also the course director and instructor for BASIC and Beyond BASIC Mechanical Ventilation and Nephrology courses in Australia and overseas.

Jason Dowling Received the B.App.Sc. and B.Comp.(Hons) degrees from Monash University, Victoria, Australia in 1998 and 2002; and the Ph.D. degree in electronic systems engineering (computer and human vision) from Queensland University of Technology, Australia in 2007. From 2007 to 2010, he was a medical image analysis Postdoctoral Fellow working with the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Since 2010, he has been a CSIRO Research Scientist working on a range of medical imaging projects. He is the author of more than 100 articles, book chapters and edited books. His research interests include image segmentation, registration, reconstruction, and clinical translation of software as a medical device. Associate Professor Dowling leads the CSIRO medical image analysis team which works closely with clinicians and industry partners to automatically analyse and report information from 3D medical images (including MRI, CT, PET, x-ray, and ultrasound) for disease diagnosis and monitoring, treatment planning, and treatment delivery. He is an expert member of ISO/IEC JTC 1/AWG 12.

Girija Chetty (Senior Member, IEEE) Received the bachelor's and master's degrees in electrical in engineering and computer science in India and the Ph.D. degree in information sciences and engineering in Australia. She has more than 35 years of experience in industry, research, and teaching from universities and research and development organizations in India and Australia. She has held several leadership positions, including the Head of Software Engineering and Computer Science, the Program Director of ITS Courses, and the Course Director of Master of Computing and Information Technology Courses. She is currently a Full Professor in computing and information technology with the School of Information Technology and Systems, University of Canberra, Australia, and her research interests include the area of multimodal systems, computer vision, 1111 pattern recognition, data mining, and medical image computing. She is an award winning academic with Vice Chancellors Excellence award in HDR/Doctoral Supervision at University of Canberra, and outstanding excellence award in Faculty of SciTech recently.

Davide Fontanarosa Professor Davide Fontanarosa is a physicist currently working in the fields of ultrasound imaging and medical physics. He worked in one of the top institutions for radiation therapy (MAASTRO Clinic, in the Netherlands) and in one of the largest industrial research laboratories in the world, Philips Research, as Senior Scientist. Then he moved to Queensland University of Technology (Brisbane, Australia) where he is currently doing research in several fields related to ultrasound, imaging techniques and radiation therapy.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

