AI-Driven Pulmonary Congestion Assessment for Lung Ultrasound via Segmentation-Guided Transformers

Fahimeh Fooladgar[1], Deepa Krishnaswamy[3], Tamas Ungi[2], Viet Dinh[3], Mike Jin[4], Matheus Alves[5,3], Caroline Schissel[6], Shreyas Puducheri[5,3], Shuhei Misawa[7], Erik Duhaime[4], Stephen Hallisey[8], Alejandra Duran Mendicuti[3], Nicole Duggan[8], Purang Abolmaesumi[1], Nicholas Harrison[7], Andrew Goldsmith[6,9], and Tina Kapur[3⋆]

[1] Department of Electrical and Computer Engineering, University of British Columbia, BC, Canada
[2] School of Computing, Queen's University, Kingston, ON, Canada
[3] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[4] Centaur Labs, Boston, MA, USA
[5] Boston University Chobanian and Avedisian School of Medicine, Boston, MA, USA
[6] Department of Emergency Medicine, Lahey Hospital, Burlington, MA, USA
[7] Department of Emergency Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
[8] Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[9] University of Massachusetts School of Medicine, Worcester, MA, USA

Abstract. Point-of-care lung ultrasound has emerged as a valuable bedside tool for diagnosing pulmonary congestion, which is often identified by the presence of hyperechoic reverberation artifacts called B-lines. However, manual B-line quantification is limited by substantial inter-operator variability and the requirement for specialized technical expertise. To address these challenges, we propose a two-stage segmentation-guided transformer framework to automate B-line severity scoring. This approach provides a consistent and reproducible assessment of pulmonary pathology directly at the point of care. On a curated clinical dataset, the proposed method significantly outperforms the baseline, improving case-level quadratic weighted kappa from 0.57 to 0.74 (difference = 0.16; 95% CI: 0.04–0.30; $p = 0.003$). The F1 score similarly increases from 0.59 to 0.75 ($p = 0.003$). The segmentation module achieves a pixel-level F1 score of 63.1% and a frame-level B-line counting F1 score of 66.7%. The model improves the accuracy and consistency of automated B-line scoring, an important step towards standardized bedside assessment of pulmonary congestion.

Keywords: Lung ultrasound · B-line detection · Segmentation

⋆ Corresponding author: tkapur@bwh.harvard.edu

1 Introduction

Point-of-care lung ultrasound (LUS) is a powerful, non-invasive tool for rapidly assessing pulmonary pathology. It offers superior sensitivity compared to traditional chest radiography for critical conditions such as pneumonia, pneumothorax, pleural effusion, pulmonary edema, and interstitial syndromes [16,24]. Despite these advantages, LUS interpretation relies heavily on advanced technical skills and is prone to subjective disagreement [5]. This is especially true for B-lines, which are hyperechoic reverberations extending from the pleural line to the bottom of the screen synchronously with lung sliding [17,14]. Since manual B-line assessment causes significant inter-operator variability [22,11], automated tools can standardize their detection and quantification.

In recent years most automated tools for B-line quantification are based on AI models and these can be broadly categorized by analysis granularity (i.e., frame, clip, or pixel-level) and by task type (classification, segmentation, or detection). For the classification of the presence/absence of B-lines, authors have utilized various convolutional neural network (CNN) approaches, including 3D CNNs [4,19,3], while others combined a CNN with a hidden Markov model [23] for clip-level classification. Others include temporal information and combined a CNN with spatial and temporal attention models, and bidirectional long short-term memory [15]. Some authors have additionally integrated multiple models with feature fusion, with the addition of classical machine learning classifiers [20]. Others utilized wavelet decomposition in their video analysis model along with an attention module [26]. However, these methods provide limited spatial interpretability and do not explicitly localize individual B-lines.

Therefore, approaches focusing on the segmentation of B-lines themselves have also been developed. One set of authors utilized crowd-sourced labels to train an Attention U-Net [2], while others proposed TransBound-UNet, combining a vision transformer with a UNet-style decoder [1]. Moving beyond the raw image analysis, other authors have analyzed the frequency components of ultrasound images, such as [7], where ultrasound frames were separated into different frequency components, enabling the network to identify B-lines while suppressing the effects of noise and unrelated anatomy. While segmentation improves spatial specificity, it typically requires dense annotations, limiting scalability. To overcome some of the issues with pixel-level delineation, authors have instead chosen to frame B-line analysis as a detection problem. One set of authors trained YOLO models for B-line bounding box detection [6], while others produced polygonal bounding boxes [21] also using YOLO [13]. Detection offers a possible compromise between the classification and segmentation of B-line approaches.

However, despite the numerous approaches for B-line analysis and quantification, several limitations still exist. Clip-level models frequently lack interpretability and spatial precision. In contrast, pixel-level approaches offer fine-grained localization, but typically overlook temporal dynamics, which restricts their capacity to understand the evolution of B-lines across frames. To address these limitations, we propose a two-stage deep learning framework that combines

frame-level B-line segmentation with temporal modeling to capture the evolution of B-lines across ultrasound clips, trained on a large-scale crowdsourced LUS dataset. The main contributions of this work are summarized as follows:

- Explicit temporal aggregation of segmentation-derived embeddings using a transformer architecture to model inter-frame dynamics for B-line analysis.
- An interpretable pipeline that localizes B-lines to provide clinically meaningful clip- and case-level severity prediction.
- A comprehensive multi-level evaluation against five state-of-the-art methods, including analyses of temporal resolution and aggregation strategies.

By combining spatial annotations with temporal modeling, our method enables more accurate and clinically useful B-line analysis compared with approaches relying solely on frame- or clip-level prediction.

2 Materials and Methods

2.1 Data

We use the MLSC-BEDLUS dataset[1], which is comprised of bedside LUS examinations from 672 emergency department patients across multiple hospitals in the Mass General Brigham network, Boston, USA. The fan-shaped beam region was isolated in 3D Slicer to extract fan corners, enabling scanline-based reconstruction and mapping to a polar coordinate system for geometry-consistent model training across diverse acquisition settings [10,2] [2]. We used a B-line-annotated subset from this dataset, created via a two-stage expert and crowdsourced process [2,12]. Acquired mainly on Mindray systems with curvilinear probes, it includes 35,923 annotated frames from 1,060 clips across 285 patients (134 female, 151 male, mean age 65.6 $\pm$ 17.4). The data are split at the patient level into training (21,170 frames, 613 clips, 169 patients), validation (4,535 frames, 140 clips, 34 patients), and test (10,218 frames, 307 clips, 82 patients) sets with no patient overlap.

2.2 Method

Problem Formulation. We consider a dataset of N lung ultrasound clips $\{\mathcal{V}^{(i)}, \{M_t^{(i)}\}, y^{(i)}\}_{i=1}^{N}$, where each clip $\mathcal{V}^{(i)} = \{I_t\}_{t=1}^{T}$ contains T grayscale frames $I_t \in \mathbb{R}^{H\times W}$ with binary segmentation masks $M_t \in \{0,1\}^{H\times W}$ indicating B-line pixels. Our objective is to predict (1) frame-wise segmentation masks $\{\widehat{M}_t\}_{t=1}^{T}$ that localize B-lines and (2) a clip-level B-line score $\widehat{y} \in \{0,1,2,3+\}$ indicating the severity of pulmonary congestion. The ground-truth clip label is defined as $y^{(i)} = \min(3, \max_t b_t^{(i)})$, where $b_t^{(i)}$ denotes the number of B-lines in frame t of clip i, computed from the ground-truth masks $\{M_t^{(i)}\}$. Following clinical convention, 3 or more B-lines denote pathologic interstitial changes (e.g. pulmonary

[1] Harvard Dataverse: DOI:10.7910/DVN/HFTV7D

[2] https://github.com/SlicerUltrasound/SlicerUltrasound

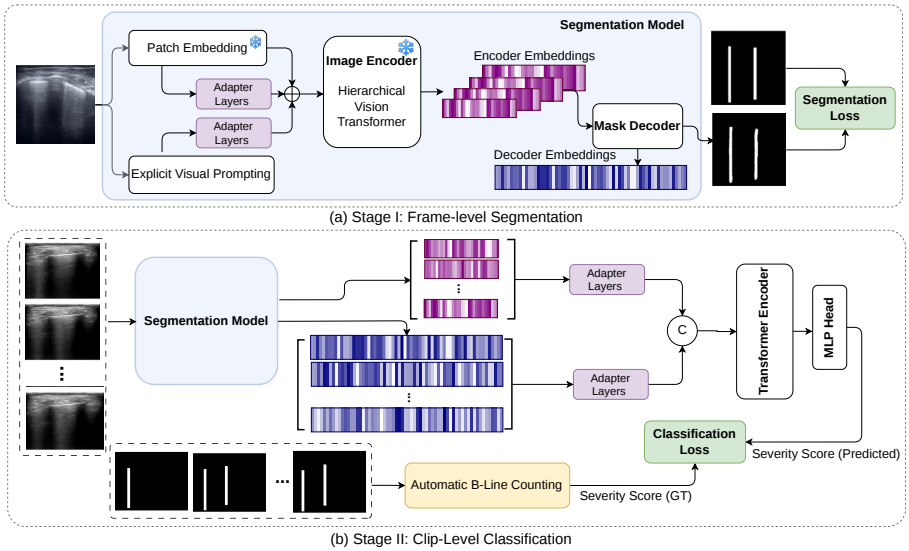


Fig. 1: End-to-end pipeline of the proposed model. A segmentation model generates per-frame B-line masks and 256-D embeddings, which are aggregated by a temporal transformer to predict clip-level B-line scores $\{0, 1, 2, 3+\}$.

edema, pneumonia, etc.) in a single zone, and multiple zones on each side, with 3 suggesting an acute process. Using the maximum count per clip parallels clinical workflows for grading of lung congestion, which is based on the most severe frame in a short sequence. Our model jointly predicts segmentation masks and clip-level B-line score class using the two-stage pipeline described below.

Stage I: Frame-Level Segmentation. We adopt SegFormer-B4 [25] with Explicit Visual Prompting (EVP) [18] as our base architecture for frame-level B-line segmentation. The encoder is a four-stage hierarchical vision transformer extracting multi-scale features. The decoder is a lightweight Multi-Layer Perceptron (MLP) that fuses multi-resolution features to produce dense per-pixel predictions (Fig. 1a). We employ the EVP to adapt the pretrained backbone, pretrained on ImageNet, to the ultrasound domain. EVP introduces two lightweight, task-specific prompt modules into the encoder while keeping encoder weights frozen:

1- Patch Embedding Prompt: Re-projects frozen patch embeddings into a task-adapted latent space to address the domain shift from natural images to ultrasound. This is done using a small adapter layer that transforms each embedding into a new representation trained for ultrasound image characteristics.

2- High-Frequency Prompt: Captures domain-specific high-frequency features by applying a 2D Fast Fourier Transform (FFT) to the input image, masking out the low-frequency band (below $\tau = 0.25$), and applying an inverse FFT. The

resulting high-frequency image is then split into patches and passed through another adapter layer to create prompts injected into each transformer stage.

The two prompts are summed, passed through a shared MLP adapter, and injected into the frozen encoder tokens. By injecting domain-relevant frequency information, EVP enables efficient adaptation to thin, high-frequency B-line patterns without fine-tuning the full encoder. The EVP modules introduce only 3.7M trainable parameters, compared to 64M in the full model.

Stage II: Clip-Level Score Classification. To estimate the clip-level B-line score, we design a transformer-based classifier that processes frame-wise embeddings extracted from the segmentation model (stage I). Each frame in the input clip is passed through a trained segmentation model, and two feature vectors are extracted: one from the final encoder stage and one from the decoder output (Fig. 1b). These are projected separately to a shared embedding space and concatenated to form a unified 512-dimensional feature vector per frame. Specifically, both the encoder and decoder outputs are passed through adaptive average pooling, linear projection, and layer normalization (Adapter Layers in Fig. 1b). The resulting vectors are concatenated into a 2×256 embedding and further projected to produce a sequence of frame tokens $\mathbf{F} = [f_1; \ldots; f_T] \in \mathbb{R}^{T \times 512}$. A learnable `[CLS]` token and positional embeddings are added before feeding the sequence into a 4-layer transformer encoder (each using four attention heads and a hidden size of 512). The `[CLS]` output is passed through a layer normalization and a final linear classifier to predict the clip-level score $\widehat{y} \in \mathbb{R}^4$.

Overall Objective. The training objective combines pixel-level segmentation and clip-level classification supervision. Given a sample $(\mathcal{V}, \{M_t\}, y)$ with T frames, the total loss is:

$$\mathcal{L} = \lambda_{\text{seg}} \cdot \Big(\tfrac{1}{T} \sum_{t=1}^{T} \mathcal{L}_{\text{seg}}(M_t, \widehat{M}_t) \Big) + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}}, \tag{1}$$

where $\mathcal{L}_{\text{cls}} = \text{CE}(\widehat{y}, y)$ is the cross-entropy loss between the predicted $\widehat{y}$ and the ground-truth clip label y. The segmentation loss is defined as a combination of binary cross-entropy (BCE) and intersection-over-union (IoU) losses:

$$\mathcal{L}_{\text{seg}} = \lambda_{\text{BCE}} \cdot \text{BCE}(M_t, \widehat{M}_t) + \lambda_{\text{IoU}} \cdot \big(1 - \text{IoU}(M_t, \widehat{M}_t)\big). \tag{2}$$

The coefficients λ_{BCE}, λ_{IoU}, λ_{seg}, and λ_{cls} balance the respective terms and are specified in the implementation details. We first train the segmentation stage independently. After convergence, the classification stage is introduced, and both stages are optimized jointly. During this phase, the learning rate of the first stage is reduced to preserve its learned representations, while the training is more significantly focused on the objective of the second stage. This joint training allows segmentation features to guide clip-level prediction via backpropagation, encouraging the model to learn features that are not only accurate for pixel-wise segmentation but also informative for assessing clip-level severity.

Implementation Details. We trained the segmentation model (Stage I) for 50 epochs using the AdamW optimizer (LR 5×10^{-4}) with a multi-step scheduler

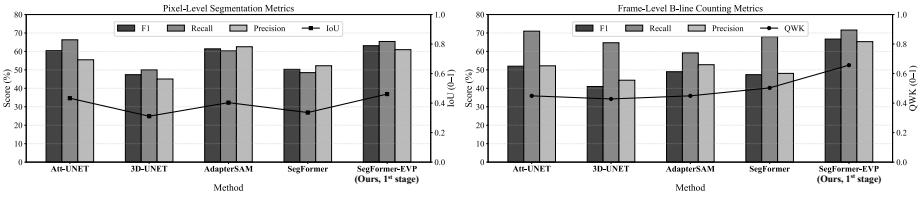


Fig. 2: Left: Pixel-wise segmentation performance, and. Right: Frame-level B-line counting results. Bar plots represent F1, recall, and precision. The line plot shows IoU for segmentation (left) and Quadratic Weighted Kappa (QWK) for counting (right), reflecting overlap quality and ordinal agreement, respectively.

(decay 0.1) and batch size of 32. The trained model was then used in Stage II, where the transformer-based classifier was trained on T-frame clips sampled with stride 1 to produce overlapping sequences for data augmentation. It was trained for another 50 epochs using AdamW (LR 10^{-5}) with a batch size of 8. The loss weights λ_{BCE}, λ_{IoU}, λ_{cls}, and λ_{seg} were set to 0.8, 0.2, 1.0, and 0.2, respectively. Hyperparameters were selected via validation tuning. Experiments were conducted on two NVIDIA Tesla V100 GPUs (16GB each). The source code is available at: `https://github.com/Fahim-F/AI-Driven-B-Line-Detection`.

Model Evaluation. We evaluate our pipeline at three levels: (1) pixel-level segmentation, (2) frame-level B-line counting, and (3) clip-/case-level classification. For segmentation and counting, SegFormer-EVP serves as the primary baseline to assess the impact of temporal aggregation. We further compare with representative architectures, including Attention U-Net [2], 3D-UNet [9], AdapterSAM [8], and the original SegFormer [25], each adapted to our dataset. To evaluate temporal aggregation with our transformer classifier, we consider two granularities: (1) *Clip-level classification*, which predicts B-line severity (0 to 3+) for each LUS clip, and (2) *Case-level classification*, which aggregates predictions across all clips from a single patient (i.e., all scanned zones from one exam). The case-level grade is computed as the average predicted clip severity, summarizing the overall severity score for the patient. For baseline methods, clip-level severity is computed from the frame with the maximum detected B-lines. In contrast, our method uses sequence-based prediction: clips with more than T frames are divided into non-overlapping sequences, and the maximum predicted severity across sequences defines the final clip-level score.

3 Results and Discussion

Pixel-Level Segmentation and Frame-Level Counting. Fig. 2-left reports pixel-wise segmentation and frame-level B-line counting. SegFormer-EVP outperforms all baselines across pixel-level metrics, achieving the highest F1 (63.1%), IoU (46.1%), recall (65.4%), and precision (61.0%). For frame-level (Fig. 2-right), it also achieves the best performance with an F1 of 66.7%, indicating improved

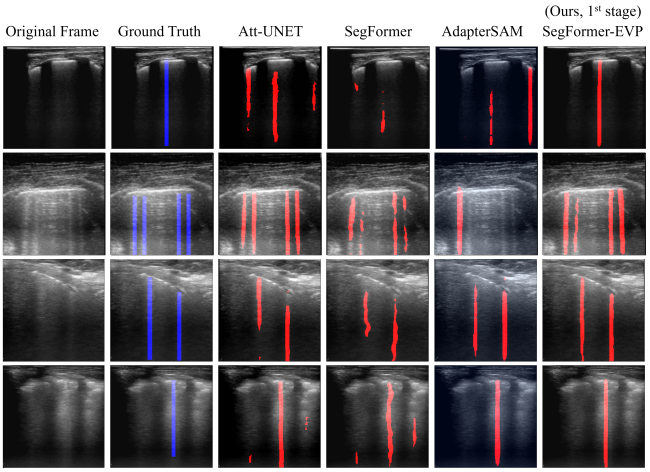


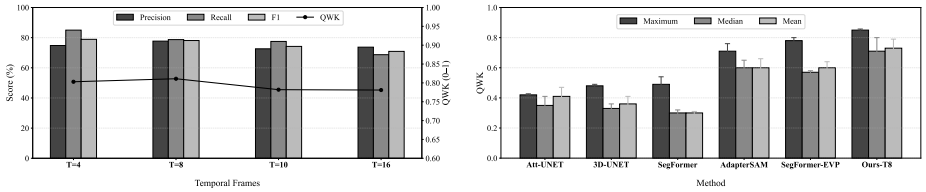
Fig. 3: Qualitative comparison of B-line segmentation on the test set across different models. Each row shows a representative frame, with columns displaying (1) the original ultrasound image, (2) ground truth B-line annotations (in blue), and (3–6) predicted segmentation masks (in red) from four different models.

Table 1: Clip-Level and Case-Level B-line Scoring: Precision, recall, F1 score, and Quadratic Weighted Kappa (QWK). "Ours-$T8$" denotes the two-stage SegFormer-EVP model with temporal aggregation over 8 frames.

	Method	**Precision**	**Recall**	**F1**	**QWK**
Clip-Level	Att-UNET [2]	54.2±3.0	53.0±5.6	40.8±5.5	0.31±0.02
	3D-UNET [9]	53.2±1.0	55.6±2.0	38.1±2.1	0.33±0.02
	SegFormer [25]	41.8±6.5	43.0±2.9	30.2±3.9	0.38±0.02
	AdapterSAM [8]	60.2±1.2	58.0±2.0	56.9±1.5	0.62±0.01
	SegFormer-EVP [18]	66.1±1.8	66.7±3.0	65.0±1.9	0.68±0.01
	Ours-$T8$ (2-stage)	**75.2±2.6**	**75.5±3.4**	**79.4±3.0**	**0.79±0.01**
Case-Level	Att-UNET [2]	65.6±3.4	50.6±2.7	46.9±3.8	0.41±0.01
	3D-UNET [9]	48.0±0.9	34.3±2.9	33.9±2.4	0.36±0.05
	SegFormer [25]	30.3±1.7	39.8±0.8	19.2±0.9	0.30±0.02
	AdapterSAM [8]	71.7±3.0	72.0±2.6	71.9±2.8	0.60±0.06
	SegFormer-EVP [18]	67.6±1.5	71.8±1.3	60.1±1.5	0.57±0.04
	Ours-$T8$ (2-stage)	**79.1±1.6**	**78.3±2.4**	**74.1±2.5**	**0.71±0.06**

consistency between segmentation outputs and B-line counts. Fig. 3 presents qualitative comparisons showing that SegFormer-EVP predictions align more closely with the ground truth in B-line location, shape, and count.

Clip-Level and Case-Level Classification. Table 1 presents performance metrics at both levels. Our model outperforms all baselines, achieving the highest F1 for both clip-level (79.4%) and case-level (74.1%) when using T8 model.



(a) Temporal resolution analysis. (b) Aggregation strategy comparison.

Fig. 4: Left: Clip-level performance across different temporal frames. Right: Case-level performance under different aggregation methods.

Overall, the results confirm the benefit of combining spatial segmentation with temporal aggregation. Unlike max-pooling aggregation, which only retains the strongest frame-level response, the proposed transformer-based aggregation incorporates contextual information across neighboring frames to improve temporal consistency and robustness against noisy frame-level predictions and transient segmentation errors. Statistical comparisons were performed against SegFormer-EVP as a baseline to isolate the effect of temporal aggregation. Case-level QWK was used as the primary metric, with F1 as a secondary metric. Ninety-five percent confidence intervals were computed using patient-level bootstrap resampling, and paired Wilcoxon signed-rank tests were used for significance testing. Our method achieved higher case-level agreement with ground truth, with a QWK of 0.74 compared with 0.57 for the baseline (difference = 0.16; 95% CI: $0.04 - 0.30$; $p = 0.003$). Secondary analysis using F1 score showed consistent improvements (0.75 vs 0.59; difference = 0.17; 95% CI: $0.04 - 0.34$; $p = 0.003$).

Ablation Studies. We evaluated the impact of temporal window size ($T = 4, 8, 10, 16$) on performance (Fig. 4a). Short windows (4-8 frames) achieved better results, with $T = 8$ yielding the highest QWK and F1. Performance is relatively consistent across short windows but decreases slightly for longer sequences, suggesting no benefit beyond 8 frames. To examine case-level aggregation, we compared mean, median, and maximum strategies. Mean reflects overall pulmonary congestion across zones, while the median provides a robust estimate of typical zone severity with reduced sensitivity to extreme clips. In contrast, maximum emphasizes worst-zone severity. As shown in Fig. 4b, mean and median yield similar results, while maximum generally produces higher QWK. Spearman correlation for our model shows $\rho = 0.97$ between mean and median and $\rho = 0.73$ between mean and maximum, indicating consistent patient rankings for mean and median and greater variability with maximum aggregation.

Limitations and future work. Although our model provides automated B-line quantification in LUS for consistent assessment across patients and clinical settings, it has several limitations. All exams originated from a single healthcare network and were collected retrospectively; therefore, exams exhibited variability in scanned lung zones and clip duration. Future work will focus on improving data quality and generalizability. Prospective and longitudinal studies with

standardized acquisition protocols will be essential to reduce variability and ensure consistent lung zone coverage. Expanding data collection across multiple healthcare networks will support validation in diverse clinical settings. Methodologically, semi-supervised and self-supervised learning may reduce reliance on dense frame-level annotations, particularly when combined with scalable strategies such as crowdsourcing. Finally, incorporating uncertainty estimation could further enhance robustness and interpretability in real-world clinical deployment.

4 Conclusion

In this work, we introduced a segmentation-guided transformer model that integrates spatial and temporal information for B-line severity classification in lung ultrasound clips. The model significantly outperforms prior methods at the pixel, frame, clip, and case levels, demonstrating that combining fine-grained segmentation with temporal context improves accuracy and consistency of pulmonary congestion assessment. Unlike previous clip-level approaches that rely on heuristic aggregation or limited temporal modeling, our method learns to attend to clinically relevant frames within each clip. Notably, our use of explicit visual prompting enables domain adaptation to ultrasound, achieving state-of-the-art results without full backbone fine-tuning.

Acknowledgments. This work was supported by NIH Grants R21EB034075 and R01EB035679, and the Massachusetts Life Sciences Bits to Bytes Award.

Disclosure of Interests. Andrew Goldsmith is a consultant for Ultrasight. Fahimeh Fooladgar is a consultant for Enchannel Medical Inc. Tamas Ungi is an employee of Claronav. These affiliations did not influence the study design, data handling, interpretation, or submission decision. All other authors declare no competing interests.

References

1. Abbasi, S., Wahd, A.S., Ghosh, S., Ezzelarab, M., Panicker, M., Chen, Y.T., Jaremko, J.L., Hareendranathan, A.: Improved A-line and B-line detection in lung ultrasound using deep learning with boundary-aware dice loss. Bioengineering **12**(3), 311 (2025)
2. Asgari-Targhi, A., Ungi, T., Jin, M., Harrison, N., Duggan, N., Duhaime, E., Goldsmith, A., Kapur, T.: Can crowdsourced annotations improve AI-based congestion scoring for bedside lung ultrasound? In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 580–590. Springer (2024)
3. Baloescu, C., Chen, A., Varasteh, A., Hall, J., Toporek, G., Patil, S., McNamara, R.L., Raju, B., Moore, C.L.: Deep-learning generated B-line score mirrors clinical progression of disease for patients with heart failure. The Ultrasound Journal **16**(1), 42 (2024)
4. Baloescu, C., Toporek, G., Kim, S., McNamara, K., Liu, R., Shaw, M.M., McNamara, R.L., Raju, B.I., Moore, C.L.: Automated lung ultrasound B-line assessment using a deep learning algorithm. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **67**(11), 2312–2320 (2020)

5. Boero, E., Gargani, L., Schreiber, A., Rovida, S., Martinelli, G., Maggiore, S.M., Urso, F., Camporesi, A., Tullio, A., Lombardi, F.A., et al.: Lung ultrasound among expert operator's: Scoring and inter-rater reliability analysis (lesson study) a secondary cows study analysis from italus group. Journal of Anesthesia, Analgesia and Critical Care **4**(1), 50 (2024)
6. Bottino, A., Botrugno, C., Casciaro, E., Conversano, F., Lay-Ekuakille, A., Lobardi, F.A., Morello, R., Pisani, P., Vetrugno, L., Casciaro, S.: Automatic approach for B-lines detection in lung ultrasound images using you only look once algorithm. Journal of Ultrasound **28**(4), 985–92 (2025)
7. Bottino, A., Botrugno, C., Conversano, Francesco, D., Koyazo, J.T., Pisani, P., Morello, R., Lay-Ekuakille, A., Sergio, C.: Variational mode decomposition for B-lines segmentation in lung ultrasound images: An improved approach. IEEE Sensors Letters (2025)
8. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: SAM-Adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023)
9. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 424–432. Springer (2016)
10. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al.: 3D slicer as an image computing platform for the quantitative imaging network. Magnetic resonance imaging **30**(9), 1323–1341 (2012)
11. Gullett, J., Donnelly, J.P., Sinert, R., Hosek, B., Fuller, D., Hill, H., Feldman, I., Galetto, G., Auster, M., Hoffmann, B.: Interobserver agreement in the evaluation of B-lines using bedside ultrasound. Journal of critical care **30**(6), 1395–1399 (2015)
12. Jin, M., Duggan, N.M., Bashyakarla, V., Mendicuti, M.A.D., Hallisey, S., Bernier, D., Stegeman, J., Duhaime, E., Kapur, T., Goldsmith, A.J.: Expert-level annotation quality achieved by gamified crowdsourcing for B-line segmentation in lung ultrasound. arXiv preprint arXiv:2312.10198 (2023)
13. Jocher, G., Chaurasia, A., Qiu, J.: Yolo by ultralytics (version 8.0. 0)[computer software] (2023)
14. Kameda, T., Kamiyama, N., Taniguchi, N.: The mechanisms underlying vertical artifacts in lung ultrasound and their proper utilization for the evaluation of cardiogenic pulmonary edema. Diagnostics **12**(2), 252 (2022)
15. Kerdegari, H., Phung, N.T.H., McBride, A., Pisani, L., Nguyen, H.V., Duong, T.B., Razavi, R., Thwaites, L., Yacoub, S., Gomez, A., et al.: B-line detection and localization in lung ultrasound videos using spatiotemporal attention. Applied Sciences **11**(24), 11697 (2021)
16. Lichtenstein, D.A., Meziere, G.A.: Relevance of lung ultrasound in the diagnosis of acute respiratory failure*: the BLUE protocol. Chest **134**(1), 117–125 (2008)
17. Lichtenstein, D.A., Meziere, G.A., Lagoueyte, J.F., Biderman, P., Goldstein, I., Gepner, A.: A-lines and B-lines: lung ultrasound as a bedside tool for predicting pulmonary artery occlusion pressure in the critically ill. Chest **136**(4), 1014–1020 (2009)
18. Liu, W., Shen, X., Pun, C.M., Cun, X.: Explicit visual prompting for low-level structure segmentations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19434–19445 (2023)

19. Lucassen, R.T., Jafari, M.H., Duggan, N.M., Jowkar, N., Mehrtash, A., Fischetti, C., Bernier, D., Prentice, K., Duhaime, E.P., Jin, M., et al.: Deep learning for detection and localization of B-lines in lung ultrasound. IEEE journal of biomedical and health informatics **27**(9), 4352–4361 (2023)
20. Moafa, K., Antico, M., Vukovic, D., Edwards, C., Canty, D., Serra, X.C., Royse, A., Royse, C., Haji, K., Dowling, J., Steffens, M., Fontanarosa, D.: Convolutional automatic identification of B-lines and interstitial syndrome in lung ultrasound images using pre-trained neural networks with feature fusion. Frontiers in Digital Health **7**, 1632376 (2026)
21. Okila, N., Katumba, A., Nakatumba-Nabende, J., Mwikirize, C., Murindanyi, S., Serugunda, J., Bugeza, S., Oriekot, A., Bossa, J., Nabawanuka, E.: Deep learning for accurate B-line detection and localization in lung ultrasound imaging. Frontiers in Artificial Intelligence **8**, 1560523 (2025)
22. Pivetta, E., Baldassa, F., Masellis, S., Bovaro, F., Lupia, E., Maule, M.M.: Sources of variability in the detection of B-lines, using lung ultrasound. Ultrasound in medicine & biology **44**(6), 1212–1216 (2018)
23. Shea, D.E., Kulhare, S., Millin, R., Laverriere, Z., Mehanian, C., Delahunt, C.B., Banik, D., Zheng, X., Zhu, M., Ji, Y., et al.: Deep learning video classification of lung ultrasound features associated with pneumonia. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3103–3112 (2023)
24. Volpicelli, G., Elbarbary, M., Blaivas, M., Lichtenstein, D.A., Mathis, G., Kirkpatrick, A.W., Melniker, L., Gargani, L., Noble, V.E., Via, G., et al.: International evidence-based recommendations for point-of-care lung ultrasound. Intensive care medicine **38**(4), 577–591 (2012)
25. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems **34**, 12077–12090 (2021)
26. Yang, K., Guo, G., Zhang, Y., Pang, L., Zheng, Z., Liu, R., Ding, J., Zhang, D., Han, J.: Frequency-aware B-line and pleural line analysis in lung ultrasound videos. IEEE Journal of Biomedical and Health Informatics (2025)