Supplementary Technical Report for : Deep Learning Video Classification of Lung Ultrasound Features Associated with Pneumonia

Daniel Shea, Sourabh Kulhare
daniel.shea@ghlabs.org, sourabh.kulhare@ghlabs.org

Appendix 1: Brief description of lung features

Pleural effusion: Pleural effusion is accumulated excess fluid in the pleural space, the region between the membranes surrounding the lung and inside the chest cavity. Pleural effusion can be caused by many conditions such as heart failure, infection and trauma. It typically manifests as chest pain and difficult breathing, and without proper management, can be life-threatening. Diagnosis of pleural effusion is not straightforward and usually requires X-ray or CT imaging. Thoracic US has been shown to enable rapid detection of pleural effusion with high accuracy, especially when radiography is not available. Studies [1] [2] have shown that US has higher accuracy than chest X-ray and is more sensitive to small volume pleural effusion (less than 200 ml, which is the limit of detection for chest X-ray). In this work, we developed a deep learning-based algorithm to automatically identify pleural effusion in US video.

Lung consolidation: Consolidation is the primary feature associated with pneumonia. A normal lung is compressible and aerated. However, when lung tissue is filled with liquid instead of air due to infection or pulmonary edema, lung consolidation occurs. Compared to normal lung, consolidation appears as a heterogeneous echotexture on US. Due to the presence of fluid/pus near the lung periphery, the margin around consolidation appears blurred and irregular. Air bronchograms are also seen with lung US similar to how they appear with chest X-ray. The consolidation algorithm is specifically trained to identify consolidations of roughly 1cm depth and larger.

B-line and Merged B-line: An artifact defined as an echogenic, coherent, and laser-like ray signal with a narrow origin in the pleural line, extending to the bottom edge of the screen, which moves synchronously with lung sliding. B-lines have been found to be a useful non-radiologic sign of extravascular lung water [3] and fibrotic lung tissue for the diagnosis of diseases such as pneumonia, interstitial lung fibrosis, and alveolar-interstitial syndrome. Thickened B-lines may fuse together to form a merged B-line (or coalescent B-line) [4]. A small number of B-lines can be found in normal lung (typically fewer than three per intercostal space). A higher density of B-lines or merged-B lines indicates loss of aeration in the lung periphery [5]. Correct interpretation of multiple B-lines/merged B-line is crucial to the diagnostic process and the clinical response. This work presents an algorithm trained to identify B-lines and merged B-lines in US video.

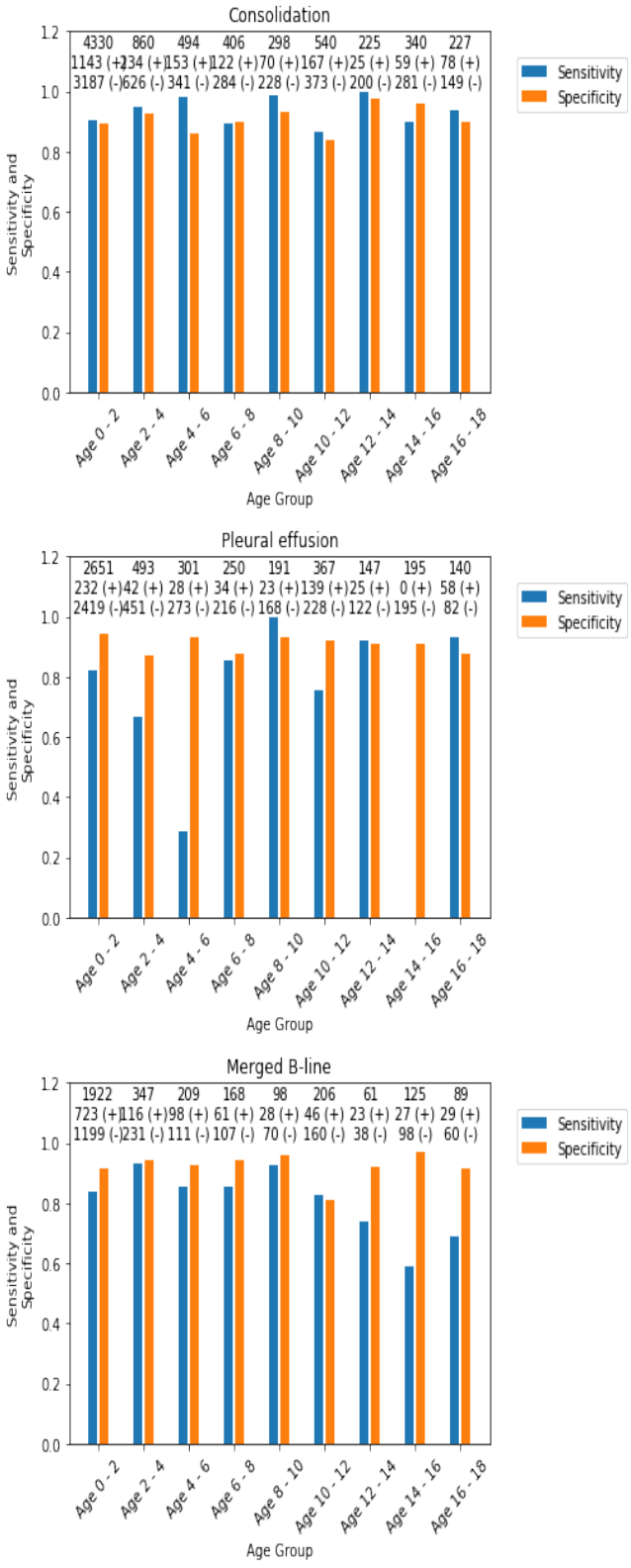

Figure 1. Age-based performance of each model. The numbers above the specificity and sensitivity bars indicate the number of videos in each age bin, the number of positive samples, and the number of negative samples. The data is only withheld test data for each of the individual models.

Appendix 2: Results of pediatric models across age groups

Pediatric patients in this work span the age range from <1 month old to 18 years old. During that age range, the human body becomes dramatically larger and the lungs imaged by US change in their characteristics. In smaller humans, more rib bones are visible in a typical lung US. This occurs because the bones are smaller and closer-together. As humans age and grow, the rib bones become more spread out resulting in a greater proportion of the US image showing lung-related features aside from the bones. This variability in the population of training images in pediatric model may cause suspicion about the model's capability to extrapolate across different age groups. Pediatric models were evaluated on different age groups across the pediatric spectrum, and results are shown in Figure 1.

The age-based performance analysis revealed relatively consistent performance across age groups with only a few age brackets experiencing reduced performance. Many of the age groups with reduced performance have relatively few positive samples, such as the pleural effusion groups for ages 2-4, ages 4-6, and ages 14-16 and the merged B-line groups for ages 10-12, 12-14, 14-16, and 16-18. This makes it difficult to evaluate the true relationship between pediatric age and model performance. However, the performance in these age bins can still be discussed with the available data. The trends presented above indicate that the merged B-line model has lower sensitivity performance on older patients in the pediatric cohort. This may indicate that an adult model may be better for older patients in the pediatric cohort. In younger patients, the definition of a merged B-line can differ when compared to older patients. In older patients, a merged B-line can be defined as multiple B-lines within a single intercostal space (the space between ribs). However, in younger patients the merged B-line is a widening B-line that is distinctly larger than a B-line. In older patients, the intercostal space is larger and thus a merged B-line can follow either the pediatric or adult definition. However, for pediatric patients only the latter definition is of importance. This difference in definition is likely incorporated in the model via an implicit bias imparted by the data annotations because the annotators labeled the US frames using both definitions for merged B-lines. However, because the "multiple B-lines in intercostal space" definition of merged B-lines has low prevalence in the pediatric data, it is possible the pediatric models did not learn that definition.

The age analysis for pleural effusion suggests that the sensitivity performance does not perform as well on younger patients. This suggests that pleural effusion is difficult to detect in younger patients. The reason behind the difficulty in identifying pleural effusion in younger patients indicates that either (i) the model doesn't generalize well to younger patients or (ii) that the sample size is just too small to make a significant judgment. In the former case, there are many reasons the model could fail to generalize for younger patients, including: a difference in the presentation of pleural effusion in younger patients, possibilities of correlation between effusion size and patient size wherein smaller effusions are more difficult to detect, or ultrasound imaging of effusions may be more challenging in younger patients.

Appendix 3: Selecting number of frames for pleural effusion

Frames processed	Validation accuracy
20	65%
30	65%
40	77%
60	84%
70	85%
80	83%
90	83%

Table 1. Effect of number of frames on validation accuracy

Processing all frames from a video leads to a high computational load. Conversely, processing fewer frames could reduce informative temporal information. Table 1 presents the effect of using different frame numbers on validation accuracy of the pediatric effusion algorithm. Validation accuracy improves significantly at 60 frames and remains approximately the same beyond 70 frames.

References

[1] N. J. Soni, R. Franco, M. I. Velez, D. Schnobrich, R. Dancel, M. I. Restrepo, and P. H. Mayo, "Ultrasound in the diagnosis and management of pleural effusions," *Journal of Hospital Medicine*, 2015. 1

[2] N. Xirouchaki, E. Magkanas, K. Vaporidi, E. Kondili, M. Plataki, A. Patrianakos, E. Akoumianaki, and D. Georgopoulos, "Lung ultrasound in critically ill patients: comparison with bedside chest radiography," *Intensive Care Medicine*, 2011. 1

[3] D. Lichtenstein, G. Meziere, P. Biderman, A. Gepner, and O. Barre, "The comet-tail artifact: an ultrasound sign of alveolar-interstitial syndrome," *American Journal of Respiratory and Critical Care Medicine*, 1997. 1

[4] G. Volpicelli, M. Elbarbary, M. Blaivas, D. A. Lichtenstein, G. Mathis, A. W. Kirkpatrick, L. Melniker, L. Gargani, V. E. Noble, G. Via, A. Dean, J. W. Tsung,

G. Soldati, R. Copetti, B. Bouhemad, A. Reissig, E. Agricola, J.-J. Rouby, C. Arbelot, A. Liteplo, A. Sargsyan, F. Silva, R. Hoppmann, R. Breitkreutz, A. Seibel, L. Neri, E. Storti, T. Petrovic, and International Liaison Committee on Lung Ultrasound (ILC-LUS) for the International Consensus Conference on Lung Ultrasound (ICC-LUS), "International evidence-based recommendations for point-of-care lung ultrasound," *Intensive Care Medicine*, 2012. 1

[5] C. F. Dietrich, G. Mathis, M. Blaivas, G. Volpicelli, A. Seibel, D. Wastl, N. S. Atkinson, X.-W. Cui, M. Fan, and D. Yi, "Lung b-line artefacts and their use," *Journal of Thoracic Disease*, 2016. 1