Automated B-Line Scoring on Thoracic Sonography

Laura J. Brattain, MS, Brian A. Telfer, PhD, Andrew S. Liteplo, MD, Vicki E. Noble, MD

With the proliferation of portable sonography and the increase in nontraditional users, there is an increased need for automated decision support to standardize results. We developed algorithms to evaluate the presence or absence of "B-lines" on thoracic sonography as a marker for interstitial fluid. Algorithm performance was compared against an average of scores from 2 expert clinical sonographers. On the set for algorithm development, 90% of the scores matched the average expert scores with differences of 1 or less. On the independent set, a perfect match was achieved. We believe that these are the first reported results in computerized B-line scoring.

Key Words—B-lines; diagnostic assistant; extravascular lung water; pulmonary edema; thoracic sonography

Received March 18, 2013, from the Massachusetts Institute of Technology Lincoln Laboratory, Lexington Massachusetts USA (L.J.B., B.A.T.); and Department of Emergency Medicine, Division of Emergency Ultrasound, Massachusetts General Hospital, Boston, Massachusetts USA (A.S.L., V.E.N.). Revision requested April 16, 2013. Revised manuscript accepted for publication May 14, 2013.

We thank Tadd Hughes, Albert Swiston, and James Williamson from the Massachusetts Institute of Technology Lincoln Laboratory for their valuable inputs on this project. This work was sponsored by the Assistant Secretary of Defense for Research and Engineering under Air Force contract FA8721-05-C-0002. The opinions, interpretations, recommendations, and conclusions are those of the authors and are not necessarily endorsed by the US Government.

Address correspondence to Laura J. Brattain, MS, Division 4, Massachusetts Institute of Technology Lincoln Laboratory, 244 Wood St, Lexington, MA 02420 USA.

E-mail: brattainl@ll.mit.edu

doi:10.7863/ultra.32.12.2185

Portable clinician-performed sonography has the potential to provide diagnostic imaging virtually anywhere needed: injury sites, ambulances, crowded emergency departments, office-based practices, rural health clinics, and war zones, to name a few. The growth of clinician-performed sonography has been mirrored by numerous reports in the literature demonstrating increased efficacy, accuracy, and comparative effectiveness when used within a focused scope of practice.[1] The devices are small, lightweight, and inexpensive and can be operated from a laptop or smart phone. The first step in realizing the full applicability of portable sonography is overcoming the time and money needed to train operators and implement a point-of-care program.

However, even after program implementation, there is still the challenge of increasing the confidence of novice users to act on their findings. Because of the proliferation of portable devices and the increase in novice users, there is a role for a computerized diagnostic assistant that can facilitate knowledge translation in these two areas. The objective metrics that a diagnostic assistant could provide would be beneficial to point-of-care sonographers since current metrics are subjective and sometimes conflicting.

Lung sonography is one application in which an objective diagnostic assistant could facilitate interpretation of images. Recent research has shown that the dynamic and static analysis of a combination of true anatomic structures and sonographic artifacts makes accurate diagnosis of many lung disorders possible, particularly when applied in emergency and critical care settings.[2] As an initial step, we consider the case of estimating a score for the presence of "B-lines," a feature of thoracic sonography that can indicate pulmonary disorders.[1] B-lines may be due to pulmonary edema, infection, contusion, or other pathologic states. Although B-lines have been shown to have good diagnostic accuracy, there is some subjectivity to their interpretation, which may limit the generalizability of thoracic sonography. Providing an objective interpretation of a B-line score would facilitate the use and reproducibility of thoracic sonography and represents an initial step toward a diagnostic assistant. We believe that this study is the first reported work in the area of computerized B-lines scoring, or a diagnostic assistant for thoracic sonography. It is important to note, however, that because B-lines are artifacts caused by physiologic changes in the lung parenchyma, they are somewhat subject to machine settings and signal processing. Standardization of postprocessing would be required if a computerized diagnostic assistant is going to be widely used in a reproducible and reliable way. This study, however, assessed the feasibility of diagnostic assistants to reliably quantify B-lines in one sample set of clips from one machine.

Materials and Methods

Patients

The patients in this study were part of a blinded prospective study of dyspnea.[3] They presented to the Massachusetts General Hospital emergency department with dyspnea and were older than 18 years and able to consent. All the patients selected were able to have thoracic sonography performed on them. There was no randomization of groups. Patients' treatment and disposition were determined by the treating emergency physician. The sonographic data were collected under approval of the Partners Human Research Committee. All patients provided written, informed consent. The use of the data for algorithm development and testing was approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

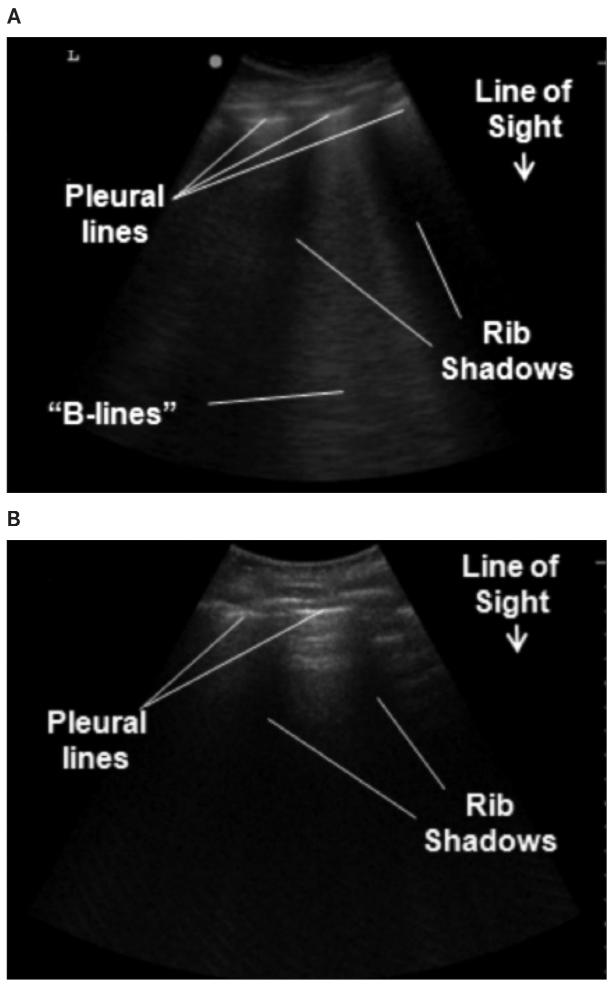
Image Acquisition

The sonograms used in this study were acquired with a MicroMaxx 2-dimensional ultrasound machine (SonoSite, Inc, Bothell, WA) equipped with an abdominal probe (60-mm broadband curvilinear probe with a frequency of 2–5 MHz). This system was a standard machine already in use clinically in the emergency department. During the data acquisition from each patient, both the left and right lung fields were scanned at a depth of 18 cm, with each lung field visualized in 4 thoracic zones (anterior-superior, anterior-inferior, lateral-superior, and lateral-inferior). The video clips were then saved to the ultrasound machine's hard drive.

B-Line Tutorial

Clinicians typically interpret thoracic sonography by first identifying rib shadows and the pleural interface located between and below two adjacent ribs, as shown in Figure 1. Because air is not a good transducer of sound waves, a sonogram of a normal, well-aerated lung will show few hyperechoic vertical lines deep to the pleural line. However, if there is interstitial fluid, sound wave reverberation is accentuated. Multiple closely spaced small horizontal reverberation artifacts appear as a bright white vertical line, which originates at the pleural line and extends to the edge of the sonographic window.[1] The same phenomenon is seen in radar returns from a cavity, which results in so-called delayed returns.[4]

Figure 1. Thoracic sonograms. **A**, Image with B-lines. **B**, Image without B-lines.



As more fluid collects, more B-lines appear, and they can start to coalesce, causing the sonogram to become more hyperechoic, or bright white. Research suggests that the number of B-lines correlates directly with the amount of interstitial fluid.[5] Clinicians have adopted scores to assess the severity of B-lines, as follows: 0, none; 1, physiologic; 2, mild; 3, moderate; and 4, severe. In practice, a score of either 3 or 4 tends to increase the likelihood of a diagnosis of increased interstitial fluid, whereas a score of either 0 or 1 will decrease it.

B-Line Scoring Algorithm

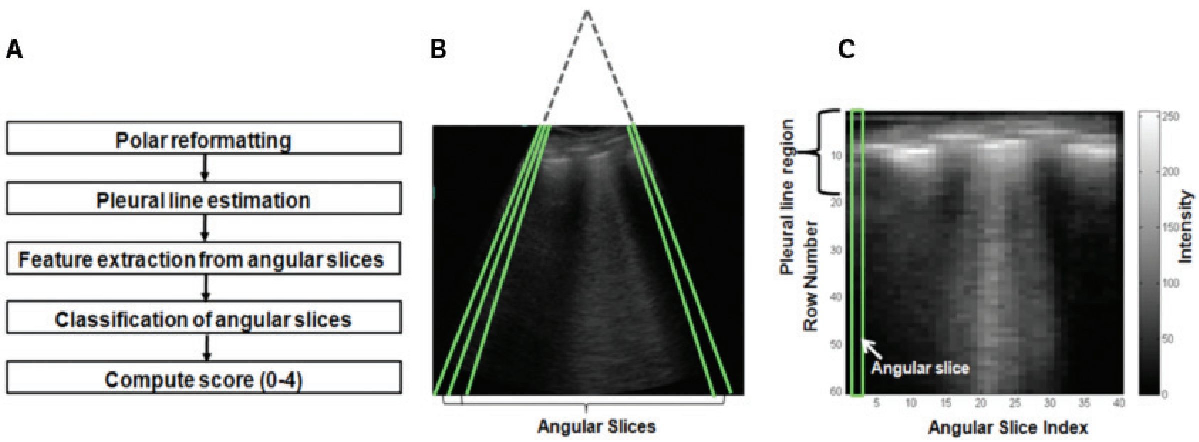
The computerized scoring algorithm was designed to identify the characteristic features of a B-line: a bright line that originates at the pleural line and extends to the edge of the sonographic window. The score was based on the number of angular slices in a video clip that contained a B-line. The algorithm needed to address individual variations (eg, the thickness of the chest wall that determines the depth of the pleural line in the image), operator variations (eg, manual changes of the gain setting), and variations in the B-line appearance due to particular statistical realizations of the closely spaced interfaces in the lungs.

The algorithm was developed on the basis of a set of 50 two-second (60 frames) video clips with B-line levels ranging from none to severe, selected by clinicians. The goal of the algorithm was to estimate a B-line score between 0 (no B-lines) to 4 (severe), as used by clinicians. Each clip was scored by a clinician using the scores described above. Scoring physicians were emergency physicians with ultrasound fellowship training and substantial experience with lung sonography. Examples of video frames with and without B-lines are shown in Figure 1.

The algorithm flow is shown in Figure 2A. In step 1, the portion of each video frame that corresponds to the sonographic data was extracted. The sonogram was reformatted by summing over 60 radial and 40 angular slices, resulting in a rectangular image with 60 rows and 40 columns. Each angular slice corresponded to a column in the rectangular image. We call this procedure polar reformatting (Figure 2B). Each frame in a video clip was extracted and reformatted. The reformatted image was then normalized with respect to image peaks to eliminate the effect of different gain settings during the original data acquisition. This normalization was intended to remove any scaling effects introduced by adjustment of the gain setting on the ultrasound device by the sonographer. The gain setting was not controlled during the data collection for this study and was not available to the algorithm. In the future, if the algorithm is integrated into the portable ultrasound software, then the gain setting would be accessible and would be used as part of the normalization process. This information could improve the algorithm results. In the absence of gain setting information, various normalization approaches were assessed, including measuring the intensity levels in the rib shadows, and the image peak approach was selected as the best option.

Figure 2C is the normalized image of Figure 1A after the reformatting. Note the pleural line region at the top of the view. It is characterized by bright clusters separated by dark rib shadows. B-lines, if there are any, would originate right below the pleural lines. We removed the pleural lines before B-line feature detection. Since the depth of the pleural lines can vary depending on the thickness of an individual's chest wall, in step 2, simple image-processing techniques such as thresholding and line detection were

Figure 2. Automated B-line scoring algorithm. **A**, Algorithm flow. **B**, Polar reformatting that shows the angular slices. **C**, Normalized image of Figure 1A after polar reformatting. Note the pleural line regions at the top of the view.

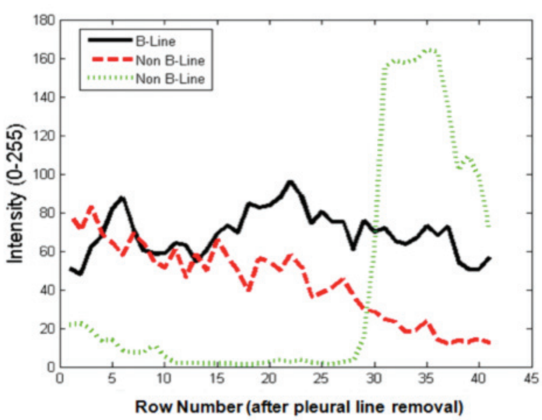


used to estimate the location of pleural lines so they could be excluded from the subsequent B-line detection steps.

In step 3, 5 features were extracted from each angular slice starting from a row that is nominally distal from the pleural lines estimated from step 2. The features were chosen to detect columns that have high, uniform intensities starting from the pleural line and that extend to the most distal point in the image. Figure 3 shows the intensity differences between a detected B-line and non–B-lines, which may have a constant low intensity or start with a very similar intensity profile as that of a B-line but taper to low intensity before reaching the end of the window (red line) or vary in other ways (green line). The red line in Figure 3 demonstrates that intensity profiles of a B-line and a non–B-line can be similar. A comprehensive feature set is needed for successful B-line detection.

The 5 features for B-line detection included the following: (1) mean of a column, (2) column length above half-maximum, (3) value of the last row of a column, (4) ratio of the value of the last row over maximum for that column, and (5) ratio of the value for the midsection of a column over maximum for that column. A B-line was detected in a particular image column if each of the 5 features exceeded a predefined threshold. An image frame was scored by summing the number of detections in that frame and then applying thresholds to the sum to map to a score. Finally, the overall score of a video clip was chosen as the maximum score over all frames in that clip. Figure 4 shows a scatterplot of 2 example features, mean and length, extracted from the same frame shown in Figure 2C. It indicates that there is some separation and identifies a list of B-line candidates. However, for a complete determination of B-lines, the other features mentioned should also be included.

Figure 3. Examples of different angular slice intensities.



Results

Initial Data Set for Algorithm Development

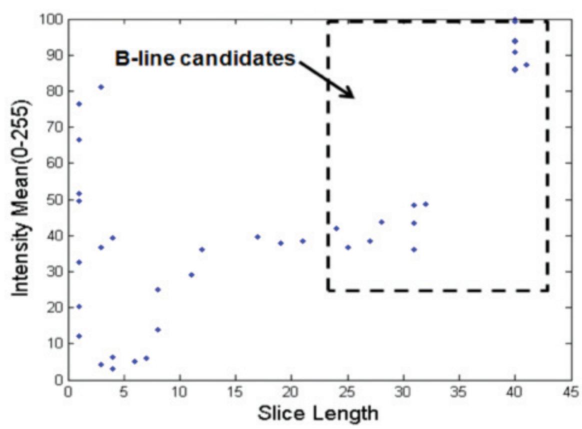
The algorithm generated scores for a set of 50 video clips used for development. The 50 clips were selected from part of a large emergency ultrasound training study, as described in "Materials and Methods." The video clips contained individual variations such as thicker chest walls in some patients and a patient's heart being in the view, and thus were judged as a challenging data set. Two expert clinicians scored each clip with scores ranging from 0 to 4. Figure 5 shows a scatterplot of the scores by the algorithm versus the average scores by the experts. The size of each point scales with the number of coincident points. The color indicates the specific number of coincident points, as shown in the key. Points falling on line B indicate perfect score matches, but any points within the boundary of lines A and C are considered clinically similar.

The absolute differences of the scores by the B-line detection algorithm and those by the experts were also computed. Differences of 1 or less were considered a match; 90% of the scores matched the expert scores. The correlation coefficient between the reference standard and the detection algorithm was 0.83.

Second Data Set for Algorithm Testing

The B-line detection algorithm was also validated with a second set of 13 video clips. This set was selected by clinicians and included negative scans without any B-lines (score of 0) and positive scans with multiple B-lines (score of 3 or 4) and was thus considered a relatively easy set for B-line detection. Our algorithm scored 100% accuracy on

Figure 4. Feature scatterplot of the image shown in Figure 3. The 2 features indicate some separation but are not sufficient for a complete determination of a B-line.



this set. The correlation coefficient between the reference standard and the detection algorithm for the test set was 1.0 (Figure 6).

Discussion

Portable sonography has the potential to provide diagnostic imaging information virtually anywhere needed. The current limiting step in realizing its full applicability is the time and money needed to train operators and implement a point-of-care sonography program. There is a need for a computerized diagnostic assistant based on sonographic video processing to aid nonexpert users. This study represents an initial step toward that goal by developing an algorithm to compute an objective score for B-lines, which can indicate the presence of extravascular lung water. Initial results are promising and demonstrate the potential in enabling portable sonographic technology.

Depending on the compression factor, the process of exporting video to a network or hard drive can affect contrast and resolution, which could alter the image quality from original to processed video. However, given that all of the clips in this study were saved in the same manner (saved directly from the machine to a portable hard drive) and that the diagnostic assistant interpretation and the expert clinical interpretation were both done on the same set of saved clips, processing of images should not affect reliability of agreement. The algorithm described herein can be applied either post hoc to video files stored on a hard drive or in real time if integrated as part of the portable ultrasound software. In the former case, care would need to be taken to store similarly high-quality video. In the latter case, it is possible that some tuning of the algorithm might be needed as part of the integration process to account for any contrast differences.

Additional work is needed to reach the goal of a real-time robust diagnostic assistant. The B-line detection algorithm can be further improved by testing against a larger and more diverse set of images (with more equivocal studies) and validated with true clinical diagnoses and clinical outcomes, in addition to the expert scores. A more sophisticated algorithm would need to integrate B-line scores over observations of different rib regions in the 4 thoracic zones. The algorithms can also be expanded to include other features, such as pleural line abnormalities and possibly consolidation patterns, as well as to detect other conditions (eg, to detect pneumothorax based on the absence of lung sliding). Finally, to provide real-time diagnostic assistance, the algorithms and augmented displays would need to be integrated with a portable sonographic device.

We expect that a diagnostic assistant for lung sonography and eventually other indications will help to increase the utility and use of portable sonography in emergency departments and hospitals and, most importantly, outside the hospital. The diagnostic assistant provides value through the following: (1) an objective measurement that can aid subjective judgments, (2) an increase in the confidence of novice users to act on their findings, and (3) a training and feedback mechanism that could help maintain and build a novice user's diagnostic skills during training. The need for these contributions can be found by considering current applications of portable sonography in the developing world and by the military.

Figure 5. Algorithm score versus experts' score for the 50 clips. The size of each point scales with the number of points that coincide. The color corresponds to a specific size, as shown in the key. Points falling on line B indicate perfect score matches, but any points within the boundary of lines A and C are considered clinically similar.

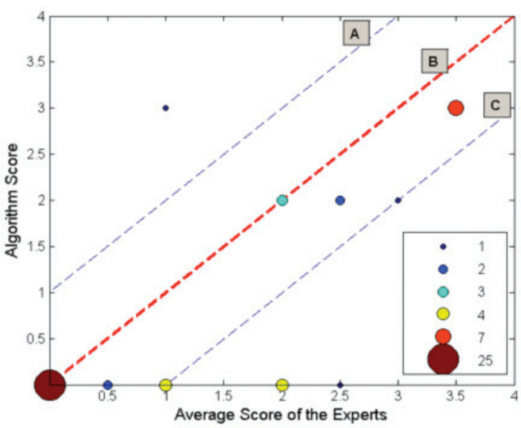
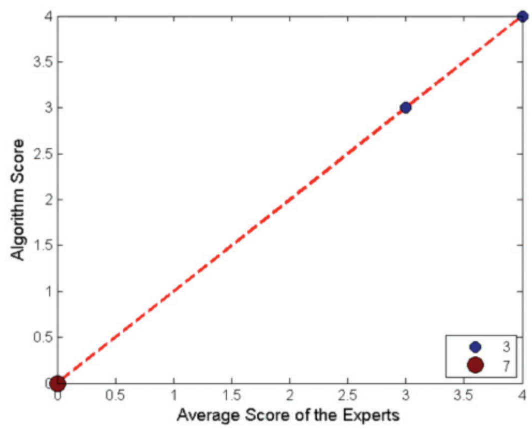


Figure 6. Algorithm score versus experts' score for the test set.



At least 9 studies of the use of sonography in the developing world have been reviewed.[6] The most prevalent diagnostic purpose varied, depending on the focus of the study, but diagnosing pneumothorax was found to be an important purpose, as part of the extended focused assessment with sonography in trauma examination. Across all the studies surveyed, considerable value of sonography was found. However, a need was identified for better quality assurance and greater uniformity. One suggestion for achieving these goals was to save sonographic video and periodically transmit the data to a sonography expert for over-reading.[7] However, there have been challenges with this solution because Internet connectivity can be insufficient or nonexistent in some remote locations. In addition, the time availability of a sonography expert could become a limiting factor. For future work, being able to objectively assess the skills of our trainees and develop a system to over-read images would greatly increase our confidence that the machine is being used appropriately and effectively in medical decision making.[7] A diagnostic assistant could substantially contribute to that goal, with increased effectiveness from real-time feedback.

A study of portable sonography by the US Special Forces reached similar conclusions.[8] Sonography was found to provide considerable value. Primary purposes were musculoskeletal (fractures and some cases of tendon and muscle body tears) and abdominal/trauma, with pneumothorax scans included as part of extended focused assessment with sonography in trauma. It was noted that, "ideally novice ultrasonographers receive real-time feedback,"[8] but that was not possible when operating in far-forward, austere environments. Again, a diagnostic assistant could provide the mechanism for real-time feedback.

Portable sonography has already greatly expanded the availability of medical imaging. A diagnostic assistant for lung sonography and eventually other indications will help realize portable sonography's full potential.

References

1 Noble VE, Nelson B. *Manual of Emergency and Critical Care Ultrasound.* 2nd ed. Cambridge, England: Cambridge University Press; 2011.
2. Volpicelli G. Lung sonography. *J Ultrasound Med* 2013; 32:165–171.
3. Liteplo AS, Marill KA, Villen T, et al. Emergency thoracic ultrasound in the differentation of the etiology of shortness of breath (ETUDES): sonographic B-lines and N-terminal pro–brain-type natriuetic peptide in diagnosing congestive heart failure. *Acad Emerg Med* 2009; 16:201–210.
4. Burrows ML. *Extended Radar Return From a Rocket Engine: A Thermal Model: MIT Lincoln Laboratory Technical Report 1123.* Lexington, MA: Massachusetts Institute of Technology; 2008.
5. Lichtenstein DA (ed). *General Ultrasound in the Critically Ill.* New York, NY: Springer; 2004.
6. Sippel S, Muruganandan K, Levine A, Shah S. Review article: use of ultrasound in the developing world. *Int J Emerg Med* 2011; 4:72.
7. Adler D, Mgalula K, Price D, Taylor O. Introduction of a portable ultrasound unit into the health services of the Lugufu Refugee Camp, Kigoma District, Tanzania. *Int J Emerg Med* 2008; 1:261–266.
8. Morgan AR, Vasios WN, Hubler DA, Benson PJ. Special operator level clinical ultrasound: an experience in application and training. *J Spec Oper Med* 2010; 10:16–21.