Butterfly™
The New Image of Health

# Clinical Value of Lung Ultrasound for B-line Detection and Quantification.


A Case Study by:

**Davinder Ramsingh, MD** | Vice President, Clinical Research & Development, Butterfly Network

**Peter Weimersheimer, MD** | Vice President, Clinical Implementation, Butterfly Network

**Harris Bergman, PhD** | Director, Deep Learning, Butterfly Network


980-23829-00 Rev A

# Introduction

Lung ultrasound is a non-invasive imaging modality that is a valuable tool in the assessment of a wide range of pulmonary diseases. This application is based on the type of artifact generated when ultrasound waves travel through tissue and interface with aerated lung pleura. In general, this is a dichotomous decision. “A-patterns” or “A-lines” are ultrasound artifacts generated when sound waves interface with normal aerated lung pleura. “B-patterns” or “B-lines” are different artifacts generated when sound waves hit an area of abnormal lung pleura in an aerated lung, often secondary to fluid or inflammation. B-lines appear as bright, vertical, hyperechoic lines that originate from the pleural surface, extend to the bottom of the ultrasound screen, and move like flashlight beams with respiration.

Both the number of B-lines and their anatomic location are suggestive of the type and severity of disease processes. One of the key advantages of lung ultrasound over chest x-ray is its superior sensitivity and specificity for detecting subtle changes in lung parenchyma associated with a variety of pathologies. For instance, chest x-ray has a sensitivity of only 60-70% for pulmonary edema.[1] When B-lines alone are used as the diagnostic criteria for this disease process, lung ultrasound has a sensitivity of over 90%.[2,3] B-line assessment is also effective in the diagnosis of interstitial lung disease. Lung ultrasound has a higher diagnostic accuracy for this pathology when compared to chest x-ray, with studies reporting sensitivities ranging from 92-100%.[1,2]

The ability to detect and quantify B-lines using lung ultrasound can provide valuable clinical information to help guide patient management. For pulmonary edema, assessing the change in frequency of B-line counts is an effective strategy for monitoring the impact of clinical intervention. It has far superior sensitivity for disease progression or remission than sequential clinical exams, imaging, and/or lab testing (NTProBNP).[3]

Lung ultrasound has several advantages over other imaging modalities. It is non-invasive and radiation-free, making it a lower-risk imaging option for patients. It is also a relatively more cost-effective, portable, and accessible option, allowing for real-time assessment of patients in various clinical settings, including the intensive care unit, emergency department, and outpatient clinics. This increased accessibility is associated with shorter times to diagnosis, shorter times to focused intervention, and more focused resource utilization.[4]

# Butterfly Ultrasound and Lung Ultrasound — Evidence-Based Care with Delay

The Butterfly iQ+ is the world's only chip-based whole-body ultrasound probe. Developed by Butterfly®, the device is portable, intuitive to use, and durable. Its single digital transducer effectively performs the same functions as three standard crystal-based probes. The Butterfly iQ+ can therefore be used for whole-body scanning and is effective in a wide variety of clinical settings. It has a lung-specific preset optimized to detect A-lines and B-lines and is an outstanding tool for lung assessment. The Butterfly iQ+ probe also has a feature within the app to rapidly change to a setting which optimizes image quality for advanced pathology such as effusions, pneumonias, and masses. With Butterfly's newly cleared lung ultrasound Auto B-line Counter tool, clinicians can now harness artificial intelligence to detect and quantify B-lines efficiently and reliably.

# The Auto B-Line Counter Tool:

Butterfly's Auto B-line Counter tool utilizes deep learning algorithms to automatically detect and quantify B-lines in lung ultrasound images, enhancing the efficiency and reliability of clinical B-line assessment. This tool couples the power of efficiently using hand-held sonography for bedside assessment with the support of AI in the assessment of respiratory disease.

The AI detects and quantifies B-lines as soon as the Auto B-line Counter tool is activated from the Actions menu. The tool will position blue overlays above the detected B-lines in real time. Discrete B-lines are indicated by a blue line and confluent B-lines are indicated by a bracket, as shown in Figure 1.

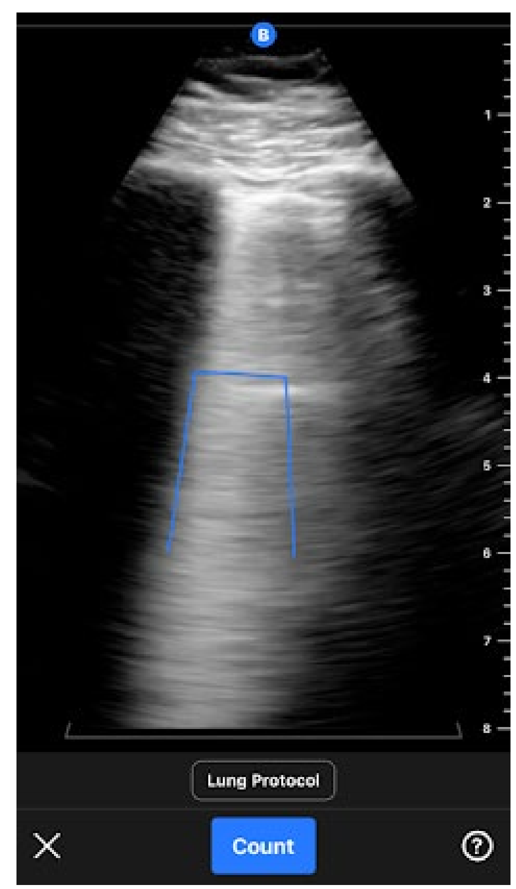


Figure 1. A confluent B-line is indicated by the bracket while scanning in real time.

Once the probe is properly positioned by the clinician with the intercostal space in the center of the screen, clicking Count begins the acquisition of a six-second clip. The AI algorithm then looks for discrete and confluent B-lines in the clip. Not only is it counting B-lines, but the AI is also performing an internal image quality check on each frame to reject counts if the image quality isn't acceptable. In moments, the tool displays the maximum number of B-lines that were observed at any one time replicating the Instant-Percent Method of B-line counting.[5]

# The Deep Learning AI that Drives the Algorithms:

The B-line counter tool uses a type of AI called deep learning. Its deep neural network is a brain-inspired algorithm that has millions of parameters arranged to convert an ultrasound frame into a list of which parts of the frame have discrete and confluent B-lines. The model predicts the location of B-lines in an image. A region is labeled as one of three classes:

1. Part of a discrete B-line
2. Part of a confluent B-line
3. Background

An innovation in the Butterfly Auto B-Lines Counting Tool is how it utilizes the instant percent counting method to assign a whole number count to confluent B-lines by the percentage of rib space occupied in addition to counting discrete B-lines — a technique that has been found to be more reliable than incumbent individual line counting methods.[6]

Another innovation is the Internal Image Quality Parameter check. This parameter approximates the fraction of sonographers who would indicate that the frame is amenable to counting B-lines. It is used internally to the algorithm and is not shown on the user interface. In pilot studies, it was shown that when the parameter was too low, there was a greater likelihood of disagreement in B-lines count between the model and expert sonographers. Consequently, only frames with image quality scores greater than or equal to a threshold are considered for the overall B-line count prediction.

# How the B-Line Counter Tool was Developed

Butterfly Network has over 2.5 million de-identified ultrasound studies in its cloud database, and thousands of studies are added each day. This volume and diverse clinical sample set allows Butterfly to approximate the natural variation of ultrasound exams being performed in clinical settings through the use of real-world evidence.

The training and validation data used in developing the B-lines Counting Tool were acquired by sampling lung studies in the cloud database and then curating the pool of studies to have an enriched dataset with a higher prevalence of cases with B-lines. Figure 2 shows the geographic distribution of the studies in Butterfly Cloud. 2,812 studies from 698 sites were used in the development of the segmentation and counting portion of the deep learning model. 1,340 studies from 295 sites were used in the development of the internal image quality parameter submodel.

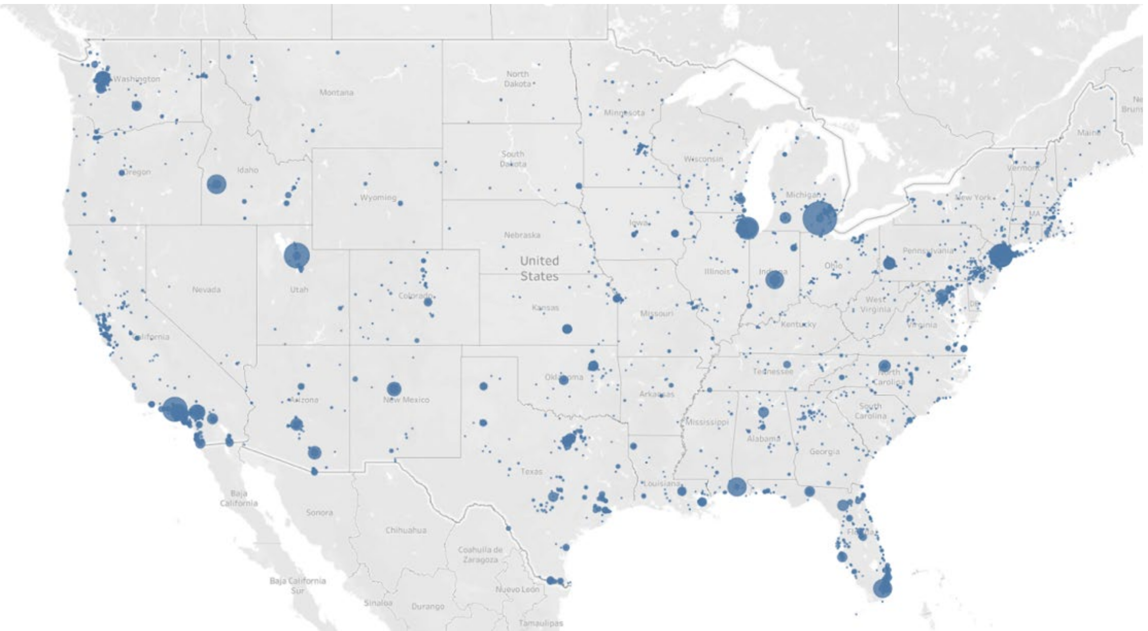


Figure 2. Geographic distribution of the studies in Butterfly Cloud.

To collect annotations for the datasets, a team of 9 board certified expert physicians with over 5 years of experience in ultrasound including a residency or specific ultrasound fellowship were recruited. These experts include Emergency Medicine, Critical Care and Radiology physicians who regularly assess patients for B-lines with ultrasound. Annotations from experts were collected via an internal, web-based tool. This web interface was provided to each annotator and allowed them to view the ultrasound cine under review, read instructions for performing the annotation, and use appropriate tools to carry out the annotation task. The annotators were assigned several tasks:

- Answer whether there are or may be B-lines in each cine
- Count the maximum number of B-lines in the cine, if any.
- Label which angular bins in a frame contained discrete and confluent B-lines. Annotators were given the previous 5 and next 5 frames to help provide context for the frame in question.
- Mark which frames in a clip are amenable to counting B-lines. This task formed the basis of the internal image quality check submodel.

## Two studies demonstrated the counter's strong performance

Two validation studies were performed to evaluate whether the performance of Auto B-line Counter was non-inferior when compared to clinician annotators (denoted as Study 1 and Study 2). The images collected for these studies represent a broad and distributed cross-section of patients, including a diverse range of B-line counts, age, gender, body mass index, ethnicity, and race.

**Study 1 Description:**

Study 1 was conducted internally at Butterfly, using a large selection of de-identified studies from the Butterfly cloud, split into development and clinical validation datasets. The objective of Study 1 was to demonstrate the Auto B-line Counter is non-inferior to clinician annotators (Ground Truth). The primary endpoint was the inter-rater correlation coefficient (ICC) between the B-line scores from the Auto B-line Counter tool and the B-line scores from the Ground Truth. The secondary endpoint was the Dice Similarity Coefficient between the centroid-paired segmentation from the Auto B-line Counter tool and the segmentation from the Ground Truth. Study 1 was a retrospective analysis of de-identified lung ultrasound studies collected during the standard usage of Butterfly probes, uploaded to the Butterfly Cloud. This data comes from the population of providers using Butterfly devices in concert with the Butterfly Cloud application in the real world. The clinical validation dataset consists of 253 de-identified six-second clips from 109 clinical sites. The data was from patients aged 22 through 90 with balanced distribution across gender.

**Study 2 Description:**

Study 2 was composed of a subset of patients from a study carried out by Chris Fung, MD, Assistant Professor of Emergency Medicine at the University of Michigan. This Auto B-line Counter Algorithm Clinical Performance Evaluation was a supplemental validation study designed to demonstrate the generalizability of the Auto B-line Counter across the relevant patient demographic categories. The primary endpoint of this study was to demonstrate the Auto B-line counter algorithm performance is non-inferior to consensus clinician interpretation (Ground Truth). The secondary objective of this study was to evaluate the algorithm's performance among diverse subgroups of age, gender, BMI/habitus, ethnicity, and race. The primary endpoint was the inter-rater correlation coefficient (ICC) between the Auto B-line Counter tool and the Ground Truth equal. Study 2 was a retrospective secondary data analysis of de-identified lung ultrasound cines and subject demographic information collected from a single site during an IRB-approved study. Data was collected from patients 22 years old or older that consented to participate in the study, and were included based on their history of admission to a general care, telemetry, or moderate care unit with clinical concerns that included pulmonary congestion. All cines were saved in the Butterfly Cloud. The data was curated by an expert clinician to cines from 97 unique subjects to ensure that there was a variety of B-lines counts present in the cines. The non-identifying subject demographic data collected included age, gender, height and weight (for BMI), ethnicity, and race; these are summarized in Table 1 below.

| Category | # of subjects |
|---|---|
| **Age (years)** | |
| *22 - 42* | *12* |
| *42 - 62* | *31* |
| *62 - 82* | *45* |
| *82 - 90* | *9* |
| **Gender** | |
| *Male* | *41* |
| *Female* | *56* |
| **BMI** | |
| *<25 kg/m2* | *27* |
| *25-30 kg/m2* | *22* |
| *30 kg/m2 or higher* | *48* |

| Category | # of subjects |
|---|---|
| **Ethnicity** | |
| *Hispanic or Latino* | *2* |
| *NOT Hispanic or Latino* | *91* |
| *Unknown / Not Reported* | *4* |
| **Race** | |
| *American Indian/Alaska Native* | *1* |
| *Black or African American* | *22* |
| *White* | *73* |
| *Unknown / Not Reported* | *1* |

Table 1. Demographic breakdown of Study 2, n=97

## B-line Visualization (aka B-line Segmentation) Performance:

Using Study 1 only, the degree of overlap in localizing the position of B-lines was assessed using the Dice Similarity Coefficient (DSC) between the centroid-paired segmentation from the Auto B-line Counter tool and the segmentation from the Ground Truth was calculated. Ground truth for B-line segmentation was determined using 7 expert annotators. The DSC was calculated between a B-line identified by the tool and a ground truth B-line that had complete or partial overlap, or abutted against one another with no overlap. As shown in Table 3, study 1 exceeded the performance goal of demonstrating the DSC was equal or larger than 0.52. The performance target was derived from published literature.

| | Acceptance Criteria | DSC | 95% CI |
|---|---|---|---|
| **Study 1 Results** | DSC $\geq$ 0.52 | 0.82 | [0.78, 0.876] |

able 3. Dice Similarity Coefficient (DSC) values from the B-lines localization study.

# Technical Summary

The Butterfly Network Auto B-Line Counter is an AI-based tool that counts the maximum number of B-lines visible in a six-second ultrasound cine. The tool demonstrated strong agreement with expert annotators in analytical and clinical validation studies.

# Benefits of a Lung Ultrasound Auto B-line Counter

The benefits of a lung ultrasound B-line counter include:

- **Improved efficiency:** The Auto B-line counter tool enables rapid and automated assessment of B-lines on lung ultrasound images, reducing the need for manual quantification by clinicians. This improves the efficiency of lung ultrasound exams, enabling clinicians to perform more exams in less time.

- **Improved precision:** The Auto B-line counter tool uses deep learning algorithms to detect and quantify B-lines, which can improve the consistency and reproducibility of the assessment.

- **Cost-effective:** Handheld ultrasound systems like the Butterfly iQ+ are more affordable and accessible than conventional cart-based models. The Butterfly Auto B-Line counter tool enhances the efficiency and accuracy of lung ultrasound exams to help improve cost-effective patient care and reduce healthcare costs.

- **Improved reliability:** The Auto B-line counter tool provides a standardized assessment of B-lines. This is especially valuable in monitoring the effectiveness of a medical intervention.

- **Increased objectivity:** The Auto B-line counter tool provides an objective assessment of B-lines.

- **Non-invasive:** Lung ultrasound exams are non-invasive and do not expose the patient to ionizing radiation, making them a relatively safe and effective imaging modality.[7]

- **Portable:** Butterfly's ultrasound devices are ultra-portable and can be brought to the bedside, enabling clinicians to perform lung ultrasound exams in a variety of clinical settings. This can help improve access to care and reduce the need for patient transport to imaging centers.

# Conclusion

The use of the Butterfly iQ+ digital handheld ultrasound improves access to care and reduces healthcare costs. Butterfly's newly approved lung ultrasound Auto B-line Counter tool offers even more benefits for clinicians who perform lung ultrasound exams. By improving the efficiency and reliability of B-line assessment, the Auto B-line counter tool can help clinicians make more informed patient care decisions.


1. Lichtenstein DA, Mezière GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. Chest. 2008 Jul;134(1):117-25. doi: 10.1378/chest.07-2800. Epub 2008 Apr 10. Erratum in: Chest. 2013 Aug;144(2):721. PMID: 18403664; PMCID: PMC3734893.

2. Martindale JL, Wakai A, Collins SP, Levy PD, Diercks D, Hiestand BC, Fermann GJ, deSouza I, Sinert R. Diagnosing Acute Heart Failure in the Emergency Department: A Systematic Review and Meta-analysis. Acad Emerg Med. 2016 Mar;23(3):223-42. doi: 10.1111/acem.12878. Epub 2016 Feb 13. PMID: 26910112.

3. Wooten WM, Shaffer LET, Hamilton LA. Bedside Ultrasound Versus Chest Radiography for Detection of Pulmonary Edema: A Prospective Cohort Study. J Ultrasound Med. 2019 Apr;38(4):967-973. doi: 10.1002/jum.14781. Epub 2018 Oct 2. PMID: 30280401.

4. https://link.springer.com/article/10.1007/s11739-020-02284-55.

5. Anderson et al "Inter-rater reliability of quantifying pleural B-lines using multiple counting methods," J. Ultrasound Med. 2013; 32:115—120.

6. Anderson et al "Inter-rater reliability of quantifying pleural B-lines using multiple counting methods," J. Ultrasound Med. 2013; 32:115—120.

7. https://www.fda.gov/radiation-emitting-products/medical-imaging/ultrasound-imaging

1. Staub LJ, Mazzali Biscaro RR, Kaszubowski E, Maurici R. Lung Ultrasound for the Emergency Diagnosis of Pneumonia, Acute Heart Failure, and Exacerbations of Chronic Obstructive Pulmonary Disease/ Asthma in Adults: A Systematic Review and Meta-analysis. J Emerg Med. 2019;56(1):53-69.https:// pubmed.ncbi.nlm.nih.gov/30314929/

2. Pagano A, Numis FG, Visone G, et al. Lung ultrasound for diagnosis of pneumonia in emergency department. Intern Emerg Med. 2015;10(7):851-854. https://pubmed.ncbi.nlm.nih.gov/26345533/

3. Pivetta E, Goffi A, Nazerian P, et al. Lung ultrasound integrated with clinical assessment for the diagnosis of acute decompensated heart failure in the emergency department: a randomized controlled trial. Eur J Heart Fail. 2019;21(6):754-766. https://pubmed.ncbi.nlm.nih.gov/30690825/