ORIGINAL ARTICLE

Open Access

Deep-learning generated B-line score mirrors clinical progression of disease for patients with heart failure

Cristiana Baloescu[1*] , Alvin Chen[2], Alexander Varasteh[1,3], Jane Hall[4], Grzegorz Toporek[2,5], Shubham Patil[2], Robert L. McNamara[6], Balasundar Raju[2] and Christopher L. Moore[1]

Abstract

Background Ultrasound can detect fluid in the alveolar and interstitial spaces of the lung using the presence of artifacts known as B-lines. The aim of this study was to determine whether a deep learning algorithm generated B-line severity score correlated with pulmonary congestion and disease severity based on clinical assessment (as identified by composite congestion score and Rothman index) and to evaluate changes in the score with treatment. Patients suspected of congestive heart failure underwent daily ultrasonography. Eight lung zones (right and left anterior/lateral and superior/inferior) were scanned using a tablet ultrasound system with a phased-array probe. Mixed effects modeling explored the association between average B-line score and the composite congestion score, and average B-line score and Rothman index, respectively. Covariates tested included patient and exam level data (sex, age, presence of selected comorbidities, baseline sodium and hemoglobin, creatinine, vital signs, oxygen delivery amount and delivery method, diuretic dose).

Results Analysis included 110 unique subjects (3379 clips). B-line severity score was significantly associated with the composite congestion score, with a coefficient of 0.7 (95% CI 0.1–1.2 $p = 0.02$), but was not significantly associated with the Rothman index.

Conclusions Use of this technology may allow clinicians with limited ultrasound experience to determine an objective measure of B-line burden.

Keywords Point-of-care ultrasound, Lung ultrasound, B-lines, Heart failure, Artificial intelligence

*Correspondence:
Cristiana Baloescu
Cristiana.Baloescu@yale.edu
[1]Department of Emergency Medicine, Yale University School of Medicine, 464 Congress Avenue, Suite 260, New Haven, Connecticut 06519, USA
[2]Philips Research Americas, 222 Jacobs Street, Cambridge, MA 02141, USA
[3]Present address: Department of Emergency Medicine, Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, MO 63110, USA
[4]Department of Emergency Medicine, University of Washington, Seattle, WA, USA
[5]Present address: Inari Medical, One Kendall Square, Building 600/700, Suite 7-501, Cambridge, MA 02139, USA
[6]Division of Cardiology, Department of Internal Medicine, Yale University School of Medicine, PO Box 208017, New Haven, CT 06520, USA

Background

Thoracic ultrasound has emerged as a vital tool in assessing hospitalized patients experiencing shortness of breath, particularly in the initial evaluation within the Emergency Department (ED) [1–4]. Lung ultrasound, in particular, can detect alveolar interstitial syndrome (AIS), indicating the presence of fluid in the lung alveolar and interstitial spaces. AIS is characterized by the appearance of "B-lines", ring-down artifactsextending from the pleural line to the bottom of the screen thatmove with respiration [5–7].

B-lines can be seen in several different conditions such as pulmonary edema in acute heart failure (HF), pneumonia, pulmonary embolus, end stage renal disease with volume overload, acute respiratory distress syndrome, and COVID-19 [4, 7–15]. The quantity and morphology of B-lines correlate with the amount of interstitial fluid present and loss of lung aeration [16, 17]. AIS severity can be a prognostic factor in heart failure and renal failure [18–20]. B-line severity in critically ill patients has been found to be predictive of mortality, length of stay and time on the ventilator [21]. However, B-line identification and quantification can vary based on user experience [22–24].

Leveraging machine learning, specifically artificial intelligence (AI) based quantification methods, may improve diagnosis when properly employed [25–27]. AI algorithms, especially those created using deep neural networks (deep learning), are increasingly utilized in medical research to generate substantial amounts of data for large scale projects [21, 28]. Employing AI to rate B-line severity allows for rapid processing of sizable datasets and increases feasibility of large-scale research studies. Automated interpretation may also allow more reproducible measures of B-line severity and could potentially be obtained by users with less experience in ultrasound (e.g. a nurse monitoring progress of a patient with pulmonary edema) and may enhance ultrasound utility in low resource settings. Employing AI in healthcare settings can refine cardiac disease evaluation, enhancing accuracy, efficiency, and personalized care, ultimately improving patient outcomes and resource allocation [29, 30].

While some studies have shown dynamic changes in B-line severity with response to treatment, fewer have examined B-line severity evolution over the treatment course and correlation with symptomatology or severity indicators, particularly in inpatient settings [18, 19, 31]. Changes in B-line severity are anticipated to parallel the clinical trajectory, potentially serving as an additional indicator of inpatient treatment advancement and efficacy. Illness severity categorization can offer valuable insights across various clinical scenarios and guide medical decision-making. Illness severity scores typically gauge the extent of illness and mirror the intricacies of the disease progression but are also used for predictive and comparative outcome assessment, resource allocation, and care process evaluation [32]. Examples include Modified Early Warning Score, Acute Physiology and Chronic Health Evaluation, Rothman index, and others [33–35]. At present, there is no established standardized score for evaluating severity during admissions for heart failure, making the assessment of clinical severity complex [36]. While congestion is primarily diagnosed clinically, a composite congestion score has been outlined to track the advancement of congestive signs and symptoms in response to standard therapy in heart failure [37]. Less is known about how B-line severity correlates with scores of illness severity categorization and if B-line severity may augment this assessment.

The aim of this study was to determine the association between a deep-learning generated B-Line severity score and the degree of pulmonary congestion severity based on clinical assessment without ultrasound (via composite congestion score) and overall illness severity (based on Rothman index) in patients with suspicion of heart failure induced pulmonary edema, and its changes in response to inpatient treatment.

Methods

Study design

This was a prospective, observational study conducted at a large academic medical center between July 2018 and May 2019. Subjects included adult English-speaking patients who presented to the Emergency Department (ED) with dyspnea and/or hypoxia, were triaged to the high acuity section of the ED, displayed B-lines on an initial screening ultrasound, and had a diagnosis of heart failure or pulmonary edema upon admission or discharge. Additionally, subjects admitted to the heart failure floor with a diagnosis of pulmonary edema or heart failure were enrolled as a supplemental cohort. The study was approved by local institutional review board. All patients meeting inclusion criteria were approached for enrollment consecutively during defined periods and invited to participate.

Subject and exam-level data collection

The study collected subject demographic information such as age, gender, race, and ethnicity, alongside various clinical parameters recorded at the time of each study ultrasound. These included vital signs, oxygen delivery method, fraction of inspired oxygen (FiO2), patient position (bed angle), recent laboratory results (troponin I, troponin T, N-terminal pro b-type natriuretic peptide), radiological imaging results (chest X-ray, lung computer tomography, transthoracic echocardiograms), and medications relevant to heart failure and pulmonary

Table 1 Composite congestion score (scale for investigator-assessed signs and symptoms of congestion) [37]

| Signs/symptoms	0	1	2	3
Dyspnea	None	Seldom	Frequent	Continuous
Orthopnoea	None	Seldom	Frequent	Continuous
Fatigue	None	Seldom	Frequent	Continuous
JVD (cm H_2O)	≤ 6	6–9	10–15	≥ 15
Rales	None	Bases	To < 50%	To > 50%
Edema	Absent/trace	Slight	Moderate	Marked

JVD, jugular venous distension

edema (type and dosage of diuretics, type and dosage of vasodilators).

Additionally, clinical data such as fluid intake and output for the 24 h before the ultrasound were documented. Intravenous and oral loop diuretic dosages were converted to intravenous furosemide equivalents [38]. Percent FiO2 was estimated by converting the flow rate into an approximate percentage. For example, two liters of oxygen via nasal cannula was considered 24% FiO2, increasing by 4% per liter up to six liters. Non-rebreather masks or high-flow nasal cannulas were approximated as 90% FiO2. Charted FiO2 was recorded for ventilated patients. Subjects were classified as having a diagnosis of heart failure or pulmonary edema based on the presence of either diagnosis in their recorded admission or discharge diagnoses. In cases where this was unclear (non-specific diagnoses such as shortness of breath, dyspnea, acute respiratory failure, etc.), electronic medical records were reviewed by a blinded author to determine the diagnosis, independent of B-line severity. A Composite Congestion Score (CCS) was calculated for each subject at the time of each research ultrasound. Dyspnea, fatigue, orthopnea, jugular venous distension, rales, and pedal edema were evaluated and documented prospectively by the research assistant using a standardized 4-point scale (Table 1) [37]. The points were then aggregated to determine the final CCS, a metric previously outlined in the EVEREST trial [37].

Additionally, the Rothman index was also collected as a potential indicator of overall illness severity [33]. The study recorded admission and discharge diagnoses of the subjects. Readmission events and survival status were monitored for six months from the initial enrollment date. Lung ultrasound clips, clinical parameters, Rothman index and CCS were collected daily until either discharge or the 10th day of hospitalization. Data collection was timed as close to 24-hour intervals as possible.

Lung ultrasound examinations

Hospitalized subjects underwent daily 8-zone lung ultrasound examinations during their inpatient stay. Lung ultrasound exams were performed by a trained research assistant using a Philips Lumify S4-1 phased
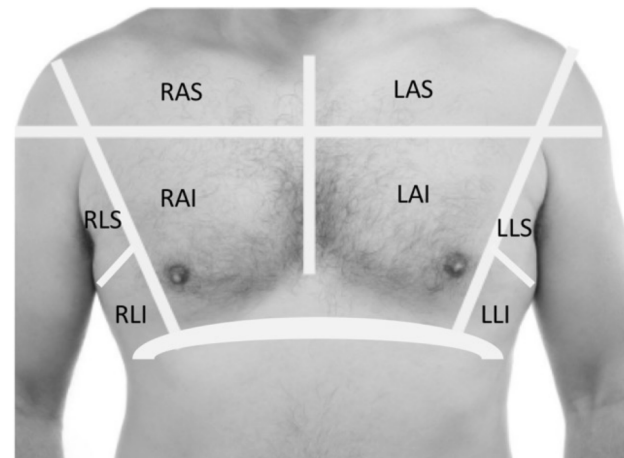
Fig. 1 Location of ultrasound interrogation of thorax for B-line presence and quantity: right anterior superior (RAS), right anterior inferior (RAI), right lateral superior (RLS), right lateral inferior (RLI), left anterior superior (LAS), left anterior inferior (LAI), left lateral superior (LLS) and left lateral inferior (LLI)

array transducer with depth set to 15 cm. Recordings of three-second clips were obtained from 8 distinct lung zones (right and left anterior superior, anterior inferior, lateral superior, and lateral inferior, Fig. 1) while the patient assumed a comfortable position, typically at approximately a 45-degree angle. For subjects enrolled in the Emergency Department (ED), ultrasound scans were repeated once on the day of enrollment, with a time gap of 2 to 5 h if the patient remained in the ED. If significant events like positive pressure ventilation, nitroglycerin drip, or diuretic administration occurred during the ED stay, and the patient was ready for transfer to an inpatient floor within 2 h, the scan was repeated sooner, 1–2 h later i. Clips were de-identified using ClipDeidentifier (www.ultrasoundoftheweek.com) for MP4 format, and DICOM Cleaner (PixelMed Publishing™) for DICOM format.

Deep-learning generated B-line score

B-line severity in each video loop was assessed using a modified version of a previously published deep learning algorithm, which rates the severity of B-lines on a scale from 0 to 4 [28, 39].
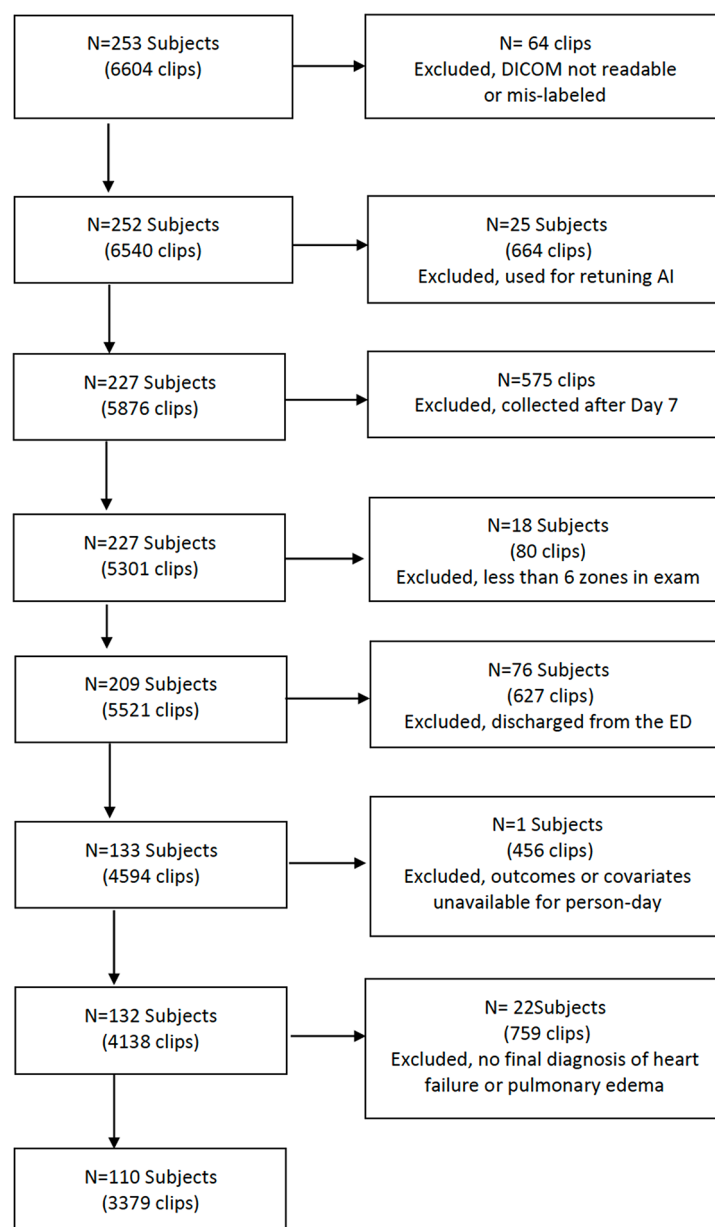
A total of 838 exams were conducted on 253 subjects, resulting in a dataset of 6,604 clips (video loops). Each video loop contained approximately 90 frames (30 frames per second). Clips that were unreadable DICOM files, mislabeled data, and those used for algorithm retuning (as described in the prior publication utilizing the same large dataset), were excluded [28]. Further excluded were exams with clips from less than 6 zones, exams from subjects discharged after ED evaluation, those lacking outcomes or model covariates available for person-day, and subjects without a diagnosis of heart failure or

pulmonary edema (Fig. 2). Final dataset contained 3379 clips from 110 subjects.

Statistical analyses

A multi-level mixed effects modeling approach was employed to explore the association between B-lines and clinical severity over time. This modeling technique allows for grouping patients to consider individual variations when dealing with longitudinal data that includes multiple measurements per patient [40]. The study evaluated the associations between these covariates and the deep-learning generated B-line score with either CCS or Rothman index. Candidate covariates encompassed demographic and clinical characteristics selected via clinician gestalt, excluding those with over 30% missing data (N-terminal pro-brain natriuretic peptide and intake-output). Reverse stepwise selection was used for model creation and eliminated the least significantly associated covariates with a stopping threshold of $p = 0.05$. Variable inflation factor was calculated to rule out for multi-collinearity among the final model covariates. No data imputation or replacement for missing data was performed. Days with missing data, either clinical or all 8-zone scores, were excluded from the analyses. All statistical analyses were performed using Stata (v.15.1, College Station, TX).


```
N=253 Subjects          →    N= 64 clips
(6604 clips)                 Excluded, DICOM not readable
                             or mis-labeled

N=252 Subjects          →    N=25 Subjects
(6540 clips)                 (664 clips)
                             Excluded, used for retuning AI

N=227 Subjects          →    N=575 clips
(5876 clips)                 Excluded, collected after Day 7

N=227 Subjects          →    N=18 Subjects
(5301 clips)                 (80 clips)
                             Excluded, less than 6 zones in exam

N=209 Subjects          →    N=76 Subjects
(5521 clips)                 (627 clips)
                             Excluded, discharged from the ED

N=133 Subjects          →    N=1 Subjects
(4594 clips)                 (456 clips)
                             Excluded, outcomes or covariates
                             unavailable for person-day

N=132 Subjects          →    N= 22Subjects
(4138 clips)                 (759 clips)
                             Excluded, no final diagnosis of heart
                             failure or pulmonary edema

N=110 Subjects
(3379 clips)
```

Fig. 2 Flow diagram of enrollment and clips utilized in the study

Table 2 Patient characteristics

N (%) or Mean (SD)	Cohort of patients with heart failure/ pulmonary edema N=110
Demographics/Enrollment	
Age	71.8 (12.9)
Female	49 (44.5)
Location of enrollment (ED)	64 (58.2)
Vital Signs	
Heart Rate	80.1 (17.6)
Systolic Blood Pressure	126.8 (22.1)
Diastolic Blood Pressure	70.1 (15.1)
Respiratory Rate	19.2 (2.4)
O2 Delivery Type	
Nasal cannula	55 (50.0)
Room air	53 (48.1)
Other	2 (1.8)
FiO_2	27.9 (12.1)
Baseline Lab studies	
Sodium	138.9 (4.5)
Hemoglobin	11.4 (2.4)
NT-ProBNP	9028.1 (13894.6)
Creatinine	1.7 (1.1)
Diuretic Type	
Combination	15 (13.6)
Loop Only	78 (70.9)
None	16 (14.5)
Diuretic Dose	99.5 (120.7)
Findings/Outcomes	
Baseline Composite Congestion Score	8.7 (3.2)
Baseline Rothman Score	61.1 (15.4)
Readmission (non-procedural) within 90 days	44 (40%)

Results

The primary study cohort, comprising 110 subjects (423 patient-days, 3379 clips), was analyzed to explore the association between clinical severity (represented by CCS and Rothman index) and deep-learning generated

Table 3 Mixed effects modeling investigating association between composite congestion score and B-line severity

| Composite Congestion Score	Coefficient	95% CI
B-line severity AI Rating	0.7	0.1 to 1.2
Exam Day (1–7)	-0.6	-0.7 to -0.5
Total loop diuretic dose	0.001	-0.001 to 0.002
Respiratory Rate	0.1	-0.02 to 0.16
Baseline Hemoglobin	-0.3	-0.5 to -0.1
O2 delivery type: room air	-2.4	-3.1 to -1.7
FiO2	-0.02	-0.06 to 0.02
Baseline Sodium	-0.1	-0.2 to -0.03

B-line severity The average age of subjects was 72 years old (±13 years), and 45% were female (Table 2). The average 8-zone B-line score over hospitalization duration was down trending, as displayed in Fig. 3a.

A weak unadjusted association was observed between composite congestion score and deep-learning generated B-line severity (Fig. 3b). Adjustment via the mixed effects regression model demonstrated a statistically significant association between the deep-learning generated B-line score and CCS, with a coefficient 0.7 (95% CI 0.1 to 1.2, $p=0.02$, Table 3). The other covariates included in the final model were total loop diuretic dose on day of exam, respiratory rate, baseline hemoglobin and sodium, FiO2, and lack of supplemental oxygen delivery (room air oxygen delivery type).

There was no significant association between Rothman index and deep-learning generated B-line severity in either unadjusted or adjusted analyses (Table 4).

Discussion

This study contributes to the body of evidence indicating that B-lines are dynamic artifacts, and that B-line severity decreases over time as congestion symptoms decrease. We have demonstrated that an AI scoring system developed using deep learning correlates with changes in clinical severity of pulmonary congestion. AI generated B-line
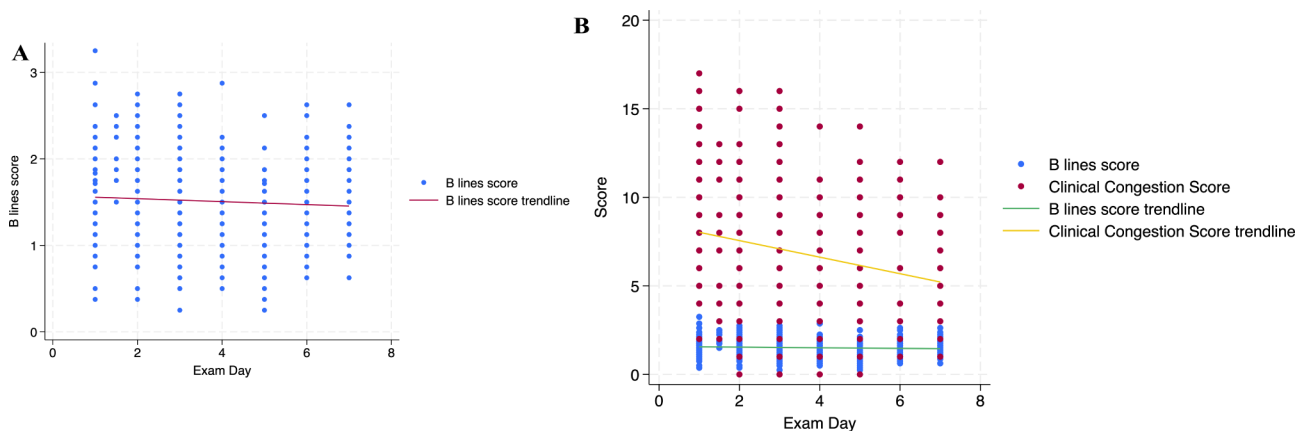
Fig. 3 **a** Eight-zone average daily B-line score trendline over hospitalization days. **b** B-line score and Composite Congestion score trendlines

Baloescu *et al. The Ultrasound Journal*   (2024) 16:42

Page 6 of 9

Table 4 Mixed effects modeling investigating association between Rothman Index and B-line severity

Rothman Index	Coefficient	95% CI
B-line severity AI Rating	-0.5	-2.5 to 1.6
Exam Day	-0.2	-0.7 to 0.3
Comorbidity – Afib	-3.2	-7.3 to 0.9
Comorbidity – COPD	-3.8	-8.7 to 1.1
Age	-0.1	-10.9 to -2.3
Female (vs. Male)	-6.6	-0.2 to 2.0
Total Loop Diuretic Dose	-0.004	-0.01 to 0.003
Diastolic blood pressure	0.09	0.004 to 0.17
Respiratory Rate	-0.8	-1.2 to -0.5
Baseline sodium	0.5	0.02 to 1.0
Baseline hemoglobin	1.4	0.5 to 2.2
FiO2	-0.1	-0.2 to 0.03

severity in our study was significantly associated with symptom burden as measured by CCS, though not associated with the Rothman index.

The decision to use CCS was informed by a study by Ambrosy et al., which investigated the progression of congestive signs and symptoms in response to standard therapy in a large contemporary cohort of patients hospitalized for worsening heart failure with reduced ejection fraction [37]. The study showed significant improvement in CCS distribution from baseline to discharge, with a simultaneous reduction in CCS correlating with sustained body weight loss [37]. Patients with higher CCS were also more frequently categorized as New York Heart Association functional class IV [37]. These findings collectively support CCS as a marker of heart failure-induced pulmonary congestion.

Although the link between CCS and B-line severity displayed statistical significance, the estimate's confidence interval is broad, suggesting a potentially weak clinical association. This may be because of some existing mismatch between congestion and B-line severity [41, 42]. B-lines have been found to persist at discharge despite lack of symptoms and clinical exam improvement: approximately 40% of patients with no rales show subclinical congestion with five or more B-lines on ultrasound at hospital discharge [43]. Additionally, in patients with renal failure, B-lines are found on ultrasound despite lack of pulmonary symptoms [20].

Another potential reason for the lack of strength of the association between CCS and B-line severity is that our study does not compare B-line severity with a fully objective assessment of clinical severity, as the CCS is partly based on symptom burden. At present, there is no established standardized and objective method for evaluating congestion in heart failure, making the assessment of clinical severity complex [36]. While chart review could have been used to assess clinical severity, conducting a daily severity review for the entire study patient population would have been highly resource-intensive for the study, and it is likely that retrospective capture of variables would yield incomplete and potentially inaccurate results.

The Rothman Index (RI, PeraHealth, Inc. Charlotte, NC, USA), an illness severity predictive model which uses continuous measurements of patient data from 26 non-static variables to measure physiologic acuity, was incorporated as a more objective comparator [33, 44]. This score is derived from vital signs, nursing assessments, laboratory findings, and cardiac rhythm, and is computed by assessing deviations from standard values, with a maximum score of 100 representing conformity to standard values. A decline in score corresponds to a deterioration in patient health. Designed to be applicable across patients regardless of diagnosis, procedure, or setting, the Rothman Index aims to offer healthcare providers a measurable, ongoing assessment of a patient's clinical condition automatically generated by the electronic medical record [33]. However, the Rothman index does consider nursing reports, which can be subjective. Our study found no significant association between B-line severity and the Rothman Index. One would expect an inverse relationship if patients were improving, but the lack of association may not be entirely surprising since the Rothman Index is designed to predict patient deterioration, not illness severity per se [33]. Unlike other early warning systems like National Early Warning Score, the Rothman Index relies on a larger number of data inputs, but these data points are given equal weight, potentially causing fluctuations in one variable to offset changes in another [33, 45]. Noise that may be introduced by additional data inputs would cancel out signal from the most critical variables. One rationale for selecting the Rothman index compared to other measures was its widespread utilization as a clinical severity metric within the medical center where the study was conducted, thus ensuring its availability for all patients. Such a score would be readily available to entire medical team easily by review of electronic medical record.

While we attempted to use available tools such as the CCS and Rothman Index, it should be understood that there is really no agreed upon gold standard, particularly one that is readily obtainable, to objectively quantify pulmonary congestion. It is conceivable that an ultrasound measure of pulmonary congestion, particularly one that is more objectively determined using AI, may ultimately outperform these existing imperfect and subjective measures.

Another limitation of the study was inclusion criteria of patients with undifferentiated dyspnea "suspected" of heart failure, alongside an admitted heart failure cohort, which resulted in some diagnostic variability. Fisease processes other than CHF and pulmonary edema cause

B-lines. Although most subjects ultimately received a discharge diagnosis of heart failure and pulmonary edema (pathologies associated with B-lines expected to improve over time), this may not have been true of patients who did not have CHF, or exhibited dual-diagnoses. For instance, a patient with lung cancer would be expected to have persistent lung findings despite their clinical course during a hospitalization, potentially biasing the results. To address this, we specifically enrolled patients with suspicion of heart failure and focused on evaluating the cohort with confirmed heart failure/pulmonary edema diagnoses to minimize the impact of this limitation in our analysis, but future studies could have stricter enrollment criteria and exclusion of other diagnoses causing fluid overload.

In patients with hypervolemia due to acute decompensated heart failure, lung ultrasound has improved sensitivity and specificity compared to chest x-ray (CXR) for pulmonary edema and pleural effusions [46–48]. In some settings, daily CXRs are common for monitoring, which can involve logistical challenges such as patient transport, costs, delays in interpretation, and radiation exposure [49]. In our study, B-line severity showed an independent association regardless of oxygenation needs and radiology results. Lung ultrasound offers a promising alternative to CXR as it is radiation sparing and can be conducted promptly at the bedside, potentially by a variety of staff members with the assistance of AI.

Using a deep-learning generated B-line score, we were able to use multi-level mixed effects modeling to test association of over twenty candidate covariates with CCS on this large dataset. Manual expert B-line scoring of >3000 clips would have been impractical. In addition to AI assisted research dataset processing, AI-assisted lung ultrasound for real-time B-line severity assessment during hospitalization or at discharge potentially holds clinical benefits. Presence of B-lines at hospital discharge of patients with heart failure indicates a five-fold increased risk for readmission or death, and presence of B-lines predicts a four-fold risk for hospitalization or death for ambulatory chronic heart failure patients [19, 50]. B-lines outperform ejection fraction as predictors for death, myocardial infarction, and heart failure progression [51, 52]. This important predictive data maymerit a different follow-up approach than similar patients without significant B-line burden and would be important information to have during a hospital admission [53]. Tracking B-line severity during hospitalization could be especially valuable for critically ill populations or those unable to communicate their symptoms. AI aids non-expert clinicians in obtaining and interpreting lung ultrasound data, and AI-based risk prediction models integrating clinical and imaging variables offer personalized assessments for heart failure management, supporting treatment decision-making [30, 54].

Conclusions

Our AI scoring system for B-line severity, generated via deep learning algorithm interpretation of lung ultrasound, was significantly associated with the composite congestion score. Use of this technology may allow clinicians with limited ultrasound experience to determine an objective measure of B-line burden. Further prospective testing of automated B-line assessment into diagnosis, prognosis, and therapy is warranted.

Declarations

References

1. Al Deeb M, Barbic S, Featherstone R, Dankoff J, Barbic D (2014) Point-of-care ultrasonography for the diagnosis of acute cardiogenic pulmonary edema in patients presenting with acute dyspnea: a systematic review and meta-analysis. Acad Emerg Med 21(8):843–852. https://doi.org/10.1111/acem.12435
2. Baid H, Vempalli N, Kumar S et al (2022) Point of care ultrasound as initial diagnostic tool in acute dyspnea patients in the emergency department of a tertiary care center: diagnostic accuracy study. Int J Emerg Med 15(1):27. https://doi.org/10.1186/s12245-022-00430-8
3. Moore CL, Copel JA (2011) Point-of-care ultrasonography. N Engl J Med 364(8):749–757. https://doi.org/10.1056/NEJMra0909487
4. Liteplo AS, Marill KA, Villen T et al (2009) Emergency thoracic ultrasound in the differentiation of the etiology of shortness of breath (ETUDES): sonographic B-lines and N-terminal pro-brain-type natriuretic peptide in diagnosing congestive heart failure. Acad Emerg Med 16(3):201–210. https://doi.org/10.1111/j.1553-2712.2008.00347.x
5. Agricola E, Bove T, Oppizzi M et al (2005) Ultrasound comet-tail images: a marker of pulmonary edema: a comparative study with wedge pressure and extravascular lung water. Chest 127(5):1690–1695. https://doi.org/10.1378/chest.127.5.1690
6. Avruch L, Cooperberg PL (1985) The ring-down artifact. J Ultrasound Med 4(1):21–28. https://doi.org/10.7863/jum.1985.4.1.21
7. Lichtenstein D, Meziere G, Biderman P, Gepner A, Barre O (1997) The comet-tail artifact. An ultrasound sign of alveolar-interstitial syndrome. Am J Respir Crit Care Med 156(5):1640–1646. https://doi.org/10.1164/ajrccm.156.5.96-07096
8. Fischer EA, Minami T, Ma IWY, Yasukawa K (2021) Lung Ultrasound for Pleural Line abnormalities, confluent B-Lines, and consolidation: Expert Reproducibility and a method of standardization. J Ultrasound Med. https://doi.org/10.1002/jum.15894
9. Soldati G (2006) Sonographic findings in pulmonary diseases. Radiol Med 111(4):507–515. https://doi.org/10.1007/s11547-006-0046-z
10. Volpicelli G, Elbarbary M, Blaivas M et al (2012) International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med 38(4):577–591. https://doi.org/10.1007/s00134-012-2513-4
11. Bouhemad B, Mongodi S, Via G, Rouquette I (2015) Ultrasound for lung monitoring of ventilated patients. Anesthesiology 122(2):437–447. https://doi.org/10.1097/ALN.0000000000000558
12. Yue Lee FC, Jenssen C, Dietrich CF (2018) A common misunderstanding in lung ultrasound: the comet tail artefact. Med Ultrason 20(3):379–384. https://doi.org/10.11152/mu-1573
13. Soldati G, Smargiassi A, Inchingolo R et al (2020) Proposal for International standardization of the Use of Lung Ultrasound for patients with COVID-19: a simple, quantitative, reproducible method. J Ultrasound Med 39(7):1413–1419. https://doi.org/10.1002/jum.15285
14. Gardelli G, Feletti F, Nanni A, Mughetti M, Piraccini A, Zompatori M (2012) Chest ultrasonography in the ICU. Respir Care 57(5):773–781. https://doi.org/10.4187/respcare.01743
15. Enghard P, Rademacher S, Nee J et al (2015) Simplified lung ultrasound protocol shows excellent prediction of extravascular lung water in ventilated intensive care patients. Crit Care 19:36. https://doi.org/10.1186/s13054-015-0756-5
16. Bouhemad B, Brisson H, Le-Guen M, Arbelot C, Lu Q, Rouby JJ (2011) Bedside ultrasound assessment of positive end-expiratory pressure-induced lung recruitment. Am J Respir Crit Care Med 183(3):341–347. https://doi.org/10.1164/rccm.201003-0369OC
17. Bitar ZI, Shamsah M, Maadarani O, Bamasood OM, Bitar AZ, Alfoudri H (2021) Lung Ultrasound and Sonographic Subpleural Consolidation in COVID-19 Pneumonia correlate with Disease Severity. Crit Care Res Pract 2021:6695033. https://doi.org/10.1155/2021/6695033
18. Palazzuoli A, Ruocco G, Beltrami M, Nuti R, Cleland JG (2018) Combined use of lung ultrasound, B-type natriuretic peptide, and echocardiography for outcome prediction in patients with acute HFrEF and HFpEF. Clin Res Cardiol 107(7):586–596. https://doi.org/10.1007/s00392-018-1221-7
19. Platz E, Merz AA, Jhund PS, Vazir A, Campbell R, McMurray JJ (2017) Dynamic changes and prognostic value of pulmonary congestion by lung ultrasound in acute and chronic heart failure: a systematic review. Eur J Heart Fail 19(9):1154–1163. https://doi.org/10.1002/ejhf.839
20. Zoccali C, Torino C, Tripepi R et al (2013) Pulmonary congestion predicts cardiac events and mortality in ESRD. J Am Soc Nephrol 24(4):639–646. https://doi.org/10.1681/ASN.2012100990

21. Tierney DM, Boland LL, Overgaard JD et al (2018) Pulmonary ultrasound scoring system for intubated critically ill patients and its association with clinical metrics and mortality: a prospective cohort study. J Clin Ultrasound 46(1):14–22. https://doi.org/10.1002/jcu.22526
22. Sustic A, Mirosevic M, Szuldrzynski K et al (2022) Inter-observer reliability for different point-of-care lung ultrasound findings in mechanically ventilated critically ill COVID-19 patients. J Clin Monit Comput 36(1):279–281. https://doi.org/10.1007/s10877-021-00726-9
23. Lerchbaumer MH, Lauryn JH, Bachmann U et al (2021) Point-of-care lung ultrasound in COVID-19 patients: inter- and intra-observer agreement in a prospective observational study. Sci Rep 11(1):10678. https://doi.org/10.1038/s41598-021-90153-2
24. Kumar A, Weng Y, Graglia S et al (2021) Interobserver Agreement of Lung Ultrasound findings of COVID-19. J Ultrasound Med 40(11):2369–2376. https://doi.org/10.1002/jum.15620
25. Russell FM, Ehrman RR, Barton A, Sarmiento E, Ottenhoff JE, Nti BK (2021) B-line quantification: comparing learners novice to lung ultrasound assisted by machine artificial intelligence technology to expert review. Ultrasound J 13(1):33. https://doi.org/10.1186/s13089-021-00234-6
26. Stewart JE, Goudie A, Mukherjee A, Dwivedi G (2021) Artificial intelligence-enhanced echocardiography in the emergency department. Emerg Med Australas 33(6):1117–1120. https://doi.org/10.1111/1742-6723.13847
27. Ahuja AS (2019) The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 7:e7702. https://doi.org/10.7717/peerj.7702
28. Baloescu C, Chen A, Varasteh A et al (2023) Two- versus 8-Zone lung Ultrasound in Heart failure: analysis of a large data set using a deep learning algorithm. J Ultrasound Med. https://doi.org/10.1002/jum.16262
29. McKee JL, Kaufman MC, Gonzalez AK et al (2023) Leveraging electronic medical record-embedded standardised electroencephalogram reporting to develop neonatal seizure prediction models: a retrospective cohort study. Lancet Digit Health 5(4):e217–e226. https://doi.org/10.1016/S2589-7500(23)00004-3
30. Ortiz J, Ghefter CG, Silva CE, Sabbatini RM (1995) One-year mortality prognosis in heart failure: a neural network approach based on echocardiographic data. J Am Coll Cardiol 26(7):1586–1593. https://doi.org/10.1016/0735-1097(95)00385-1
31. Zhu Y, Li N, Wu M et al (2022) Impact of B-lines-guided intensive heart failure management on outcome of discharged heart failure patients with residual B-lines. ESC Heart Fail 9(4):2713–2718. https://doi.org/10.1002/ehf2.13988
32. Ridley S (1998) Severity of illness scoring systems and performance appraisal. Anaesthesia 53(12):1185–1194. https://doi.org/10.1046/j.1365-2044.1998.00615.x
33. Rothman MJ, Rothman SI, Beals Jt (2013) Development and validation of a continuous measure of patient condition using the Electronic Medical Record. J Biomed Inf 46(5):837–848. https://doi.org/10.1016/j.jbi.2013.06.011
34. Subbe CP, Kruger M, Rutherford P, Gemmel L (2001) Validation of a modified early warning score in medical admissions. QJM 94(10):521–526. https://doi.org/10.1093/qjmed/94.10.521
35. Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) APACHE II: a severity of disease classification system. Crit Care Med 13(10):818–829. https://www.ncbi.nlm.nih.gov/pubmed/3928249
36. Huston JH, Ferre R, Pang PS, Chioncel O, Butler J, Collins S (2018) Optimal endpoints of Acute Heart failure therapy. Am J Ther 25(4):e465–e474. https://doi.org/10.1097/MJT.0000000000000792
37. Ambrosy AP, Pang PS, Khan S et al (2013) Clinical course and predictive value of congestion during hospitalization in patients admitted for worsening signs and symptoms of heart failure with reduced ejection fraction: findings from the EVEREST trial. Eur Heart J 34(11):835–843. https://doi.org/10.1093/eurheartj/ehs444
38. Testani JM, Brisco MA, Turner JM et al (2014) Loop diuretic efficiency: a metric of diuretic responsiveness with prognostic importance in acute decompensated heart failure. Circ Heart Fail 7(2):261–270. https://doi.org/10.1161/CIRCHEARTFAILURE.113.000895
39. Baloescu C, Toporek G, Kim S et al (2020) Automated Lung Ultrasound B-Line Assessment using a deep learning algorithm. IEEE Trans Ultrason Ferroelectr Freq Control 67(11):2312–2320. https://doi.org/10.1109/TUFFC.2020.3002249
40. Detry MA, Ma Y (2016) Analyzing repeated measurements using mixed models. JAMA 315(4):407–408. https://doi.org/10.1001/jama.2015.19394
41. Volpicelli G, Skurzak S, Boero E et al (2014) Lung ultrasound predicts well extravascular lung water but is of limited usefulness in the prediction of wedge pressure. Anesthesiology 121(2):320–327. https://doi.org/10.1097/ALN.0000000000000300

42. Imanishi J, Maeda T, Ujiro S et al (2023) Association between B-lines on lung ultrasound, invasive haemodynamics, and prognosis in acute heart failure patients. Eur Heart J Acute Cardiovasc Care 12(2):115–123. https://doi.org/10.1093/ehjacc/zuac158

43. Rivas-Lasarte M, Maestro A, Fernandez-Martinez J et al (2020) Prevalence and prognostic impact of subclinical pulmonary congestion at discharge in patients with acute heart failure. ESC Heart Fail 7(5):2621–2628. https://doi.org/10.1002/ehf2.12842

44. Gotur DB, Masud F, Paranilam J, Zimmerman JL (2020) Analysis of Rothman Index Data to Predict Postdischarge adverse events in a medical Intensive Care Unit. J Intensive Care Med 35(6):606–610. https://doi.org/10.1177/0885066618770128

45. Bilben B, Grandal L, Sovik S (2016) National early warning score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient - a prospective observational study. Scand J Trauma Resusc Emerg Med 24:80. https://doi.org/10.1186/s13049-016-0273-9

46. Maw AM, Hassanin A, Ho PM et al (2019) Diagnostic accuracy of point-of-care lung ultrasonography and chest radiography in adults with symptoms suggestive of Acute Decompensated Heart failure: a systematic review and Meta-analysis. JAMA Netw Open 2(3):e190703. https://doi.org/10.1001/jamanetworkopen.2019.0703

47. Nakao S, Vaillancourt C, Taljaard M, Nemnom MJ, Woo MY, Stiell IG (2021) Diagnostic accuracy of Lung Point-Of-Care Ultrasonography for Acute Heart failure compared with chest X-Ray study among dyspneic older patients in the Emergency Department. J Emerg Med 61(2):161–168. https://doi.org/10.1016/j.jemermed.2021.02.019

48. Yousefifard M, Baikpour M, Ghelichkhani P et al (2016) Screening performance characteristic of Ultrasonography and radiography in detection of Pleural Effusion; a Meta-analysis. Emerg (Tehran) 4(1):1–10. https://www.ncbi.nlm.nih.gov/pubmed/26862542

49. Ganapathy A, Adhikari NK, Spiegelman J, Scales DC (2012) Routine chest x-rays in intensive care units: a systematic review and meta-analysis. Crit Care 16(2):R68. https://doi.org/10.1186/cc11321

50. Li Y, Ai H, Ma N, Li P, Ren J (2022) Lung ultrasound-guided treatment for heart failure: an updated meta-analysis and trial sequential analysis. Front Cardiovasc Med 9:943633. https://doi.org/10.3389/fcvm.2022.943633

51. He J, Yi S, Zhou Y et al (2022) B-Lines by Lung Ultrasound can predict worsening Heart failure in Acute Myocardial Infarction during hospitalization and short-term Follow-Up. Front Cardiovasc Med 9:895133. https://doi.org/10.3389/fcvm.2022.895133

52. Gargani L, Pugliese NR, Frassi F et al (2021) Prognostic value of lung ultrasound in patients hospitalized for heart disease irrespective of symptoms and ejection fraction. ESC Heart Fail 8(4):2660–2669. https://doi.org/10.1002/ehf2.13206

53. Rivas-Lasarte M, Alvarez-Garcia J, Fernandez-Martinez J et al (2019) Lung ultrasound-guided treatment in ambulatory patients with heart failure: a randomized controlled clinical trial (LUS-HF study). Eur J Heart Fail 21(12):1605–1613. https://doi.org/10.1002/ejhf.1604

54. Nhat PTH, Van Hao N, Tho PV et al (2023) Clinical benefit of AI-assisted lung ultrasound in a resource-limited intensive care unit. Crit Care 27(1):257. https://doi.org/10.1186/s13054-023-04548-w

Publisher's Note