

ORIGINAL RESEARCH

Imaging

Artificial intelligence versus expert: a comparison of rapid visual inferior vena cava collapsibility assessment between POCUS experts and a deep learning algorithm

Michael Blaivas MD, MBA¹ | Srikar Adhikari MD, MPH² | Eric A. Savitsky MD³ |
 Laura N. Blaivas BS⁴ | Yiju T. Liu MD⁵

¹ Department of Emergency Medicine, St. Francis Hospital, School of Medicine, University of South Carolina, Columbia, South Carolina, USA

² Department of Emergency Medicine, School of Medicine, University of Arizona, Tucson, Arizona, USA

³ Department of Emergency Medicine, UCLA David Geffen School of Medicine, UCLA Ronald Reagan Medical Center, Los Angeles, California, USA

⁴ Department of Emergency Medicine, Harbor-UCLA Medical Center, David Geffen School of Medicine, UCLA, Los Angeles, California, USA

⁵ Michigan State University-East Lansing, East Lansing, Michigan, USA

Correspondence

Michael Blaivas, MD, MBA, FACEP, FAIUM, PO Box 769209 Roswell, GA 30076, USA.
 Email: mike@blaivas.org

Funding and support: By *JACEP Open* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

[Correction added on 14 August 2020, after first online publication: the surname of the author Yiju T. Liu is corrected from "Lui" to "Liu".]

Abstract

Objectives: We sought to create a deep learning algorithm to determine the degree of inferior vena cava (IVC) collapsibility in critically ill patients to enable novice point-of-care ultrasound (POCUS) providers.

Methods: We used publicly available long short term memory (LSTM) deep learning basic architecture that can track temporal changes and relationships in real-time video, to create an algorithm for ultrasound video analysis. The algorithm was trained on public domain IVC ultrasound videos to improve its ability to recognize changes in varied ultrasound video. A total of 220 IVC videos were used, 10% of the data was randomly used for cross correlation during training. Data were augmented through video rotation and manipulation to multiply effective training data quantity. After training, the algorithm was tested on the 50 new IVC ultrasound video obtained from public domain sources and not part of the data set used in training or cross validation. Fleiss' κ was calculated to compare level of agreement between the 3 POCUS experts and between deep learning algorithm and POCUS experts.

Results: There was very substantial agreement between the 3 POCUS experts with $\kappa = 0.65$ (95% CI = 0.49–0.81). Agreement between experts and algorithm was moderate with $\kappa = 0.45$ (95% CI = 0.33–0.56).

Conclusions: Our algorithm showed good agreement with POCUS experts in visually estimating degree of IVC collapsibility that has been shown in previously published studies to differentiate fluid responsive from fluid unresponsive septic shock patients. Such an algorithm could be adopted to run in real-time on any ultrasound machine with a video output, easing the burden on novice POCUS users by limiting their task to obtaining and maintaining a sagittal proximal IVC view and allowing the artificial intelligence make real-time determinations.

Supervising Editor: Steven G. Rothrock, MD

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *JACEP Open* published by Wiley Periodicals LLC on behalf of the American College of Emergency Physicians.

KEYWORDS

artificial intelligence, critical care, deep learning, fluid responsiveness, inferior vena cava, point-of-care ultrasound

1 | INTRODUCTION

Point-of-care ultrasound (POCUS) assessment of the inferior vena cava (IVC) has evolved over time in response to various studies either supporting or questioning its use as a non-invasive corollary for patient volume status, central venous pressure, and right atrial pressure.¹⁻³ Although initially heralded as an accurate assessment of patient volume status dating back to the 1990s, IVC ultrasound use has been appropriately challenged due to the variability of study results and inter-rater reliability challenges.^{4,5} Practical clinical experience and a careful analysis of confounding literature indicate that IVC collapsibility is most useful in its extremes, either relatively flat with significant collapse, or a plethoric state with little to no diameter variation throughout the respiratory cycle in the spontaneous breathing patient. Recent literature has supported IVC collapsibility index ([IVC expiratory diameter-IVC inspiratory diameter]/IVC expiratory diameter) use to predict critically ill patient fluid responsiveness in shock states.^{6,7} Two recent studies have indicated that a 25% collapsibility index may be adequately sensitive to differentiate between fluid responsive shock patients and fluid unresponsive ones, resulting in an ROC of 0.82 for predicting fluid responsiveness.^{6,7}

Some authors have suggested significant difference in area under the curve (AUC) results between novice sonologists performing manual measurements and calculating IVC collapsibility at the patient's bedside and the same measurements by POCUS experts made at a later time when reviewing DICOM video of these patients.⁷ Similarly, other studies have noted significant inter-rater variability between novice and expert sonologist.^{8,9} Although not surprising, this raises the question of reliability in clinical situations and also highlights the workload burden imposed on novices if they have to freeze images, move forward and backward 1 frame at a time through a cine loop to find maximal and minimal diameters of an IVC, and then carefully measure at the same location. Automation of the process holds the potential to improve inter-rater reliability and even automating the steps of performing and documenting the ultrasound examination in evaluating the IVC.¹⁰

Artificial intelligence is rapidly infiltrating modern medicine. Deep learning, a branch of artificial intelligence, is currently the most promising application for medical image analysis and interpretation. Considerable work has been performed in consultative diagnostic imaging including automatic analysis of computed tomography (CT), chest x-ray (CXR), and magnetic resonance imaging (MRI).¹¹⁻¹³ Deep learning applications can also be found in ultrasound, but have been largely limited to costly imaging platforms.^{14,15} However, applications focused on POCUS, which is directly used by clinicians in patient management including for immediate decisionmaking about fluid resuscitation

and vasopressor use, has had relatively little deep learning application growth commercially and academically. Recently, we have seen the introduction of deep learning applications in POCUS devices, including automated left ventricular ejection fraction (EF) assessment with more automation promised.¹⁶

We developed a deep learning algorithm with temporal tracking to assist in real-time video interpretation required for ultrasound assessment of IVC collapsibility and possible prediction of fluid responsiveness in critically ill patients or patient fluid status.

2 | METHODS

2.1 | Study design

This was a study of deep learning algorithm development to automatically assess whether selected ultrasound videos showed IVC collapsibility to be $\geq 25\%$. The study was Institutional Review Board-exempted with all data coming from public domain open access sources with no patient identifiers, without actual patient data being used.

2.2 | Data

Ultrasound video data of sagittal proximal IVC ultrasound examinations was obtained from public domain open access sources with all patient identifying information removed. Ultrasound video sources included anonymized image bank repositories, internet posted videos, stock videos, ultrasound vendor video and videos covering IVC evaluation, and cardiac evaluation categories. Internet search criteria included "video," "ultrasound," "IVC," "inferior vena cava," "volume status," "shock," "dehydration," and "resuscitation." Videos from the internet were downloaded using open source software called Youtube-DL. Severity of illness related to subjects of the videos was gauged from the associated vignette or history provided. Researchers specifically avoided to more closely mirroring real-life patient distributions in a clinical setting. Video data types included WMV, MP1, MP2, MOV, AVI, and MP4, and extracted single frames were all JPG. A total of 220 proximal IVC videos were imported into the training dataset. No patient identifiers were present on any of the image sources. Extracted videos included critically ill patients undergoing resuscitation as well as patients who were not critical. IVC status ranged from severely volume overloaded to severe volume depleted. No sample size or power calculations were made for this study. All reviewed and identified videos that were extracted were used for the project.

2.3 | Data manipulation and labeling

All videos, used for deep learning algorithm training, were kept in their original aspect ratio and size and sorted into 1 of 2 categories. A POCUS expert with 28 years of research, education, and clinical use experience and >200 peer reviewed research manuscript publications, ranked all videos as either collapsing at least 25%, or collapsing <25%. The assessments were made visually using video editing software MicroDicom (Sofia, Bulgaria) that allowed frame-by-frame advance and reverse viewing (Figure 1). Diameters were compared to the reference centimeter scales present on ultrasound videos for greater accuracy.

All training videos were then augmented using FFMPEG open access software. This is a common technique in deep learning when limited data are available. The total size of the training video dataset was increased 6-fold by adding copies of the original videos that were transformed by flipping them horizontally, vertically, and 45° clockwise and counterclockwise. Videos were not otherwise adjusted such as through manipulation of contrast, sharpness, and changing image quality. This approach is also helpful in preparing the algorithm for videos of IVCs that are not perfectly horizontal on the screen and can significantly increase the robustness of an algorithm for interpreting novel imaging data.

2.4 | Algorithm design

We made use of the Python programming language version 3.72 with Anaconda to manage packages and help in scripting and use of a VGG-16 convolutional neural network incorporated into a bidirectional long short term memory (LSTM) network. Code for VGG-16 is available from various public sources including github.com. VGG-16, which won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in 2014, is an early convolutional neural network using 16 layers and has been shown to be superior for ultrasound deep learning applications in prior work.¹⁷ LSTM refers to a structure that incorpo-

The Bottom Line

Inferior vena cava collapsibility (cIVC) of 25% predicts fluid responsiveness in critically ill patients. The authors of this study developed a deep learning algorithm to assist with real-time cIVC measurement during POCUS using publicly available videos. Their deep learning algorithm showed good agreement with 3 POCUS experts in determining the degree of cIVC.

rates a convolutional neural network into a structure that tracks temporal changes on ultrasound images and is used in non-medical circles to identify specific action on sports video and even predict outcomes of movements or actions. The bidirectional aspect means the video is moved forward and backward through the LSTM network. This bidirectionality helps the network to better understand the video context. We used this same basic architecture and adapted it to ultrasound video analysis.

We trained our LSTM algorithm using a PC with an 11 GB NVIDIA GeForce RTX 2080 Ti GPU and 64 GB of RAM. Researchers manipulated optimizers, learning rates, and batch size during training for optimal training times and accuracies, but avoided exploding gradients that result in training failure. The number of epochs (that is defined as 1 round of training through all of the data) was adjusted for optimal results while avoiding overfitting. Best performance was eventually obtained with 80 Epochs. Batch sizes of 15 videos ultimately proved to result in the best algorithm training performance.

2.5 | Algorithm validation and testing

The LSTM algorithm automatically performs cross validation after every epoch. Cross validation accuracy, learning, and training losses

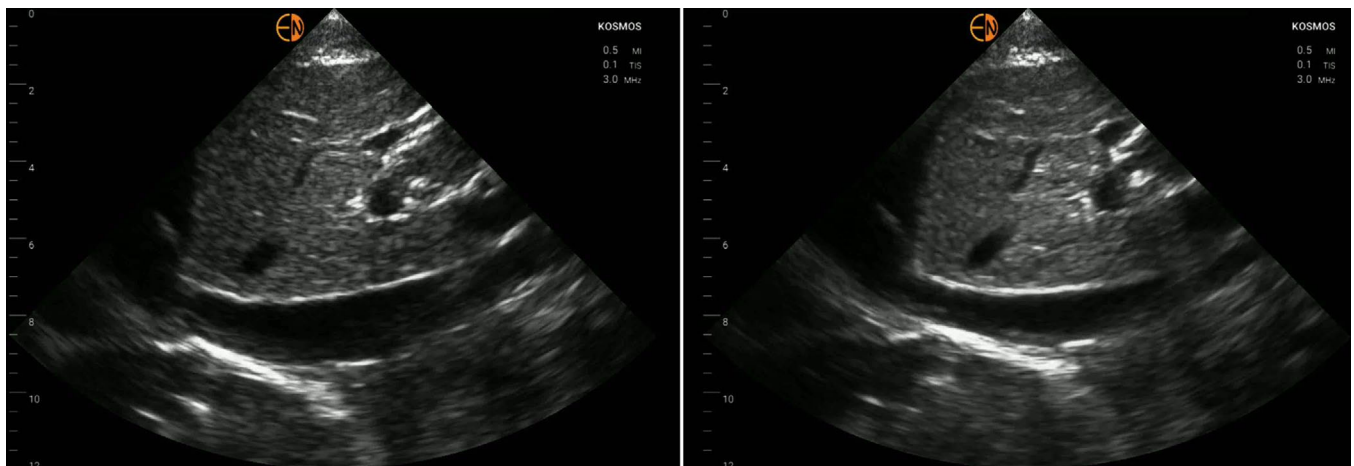


FIGURE 1 Maximal (left) and minimum (right) inferior vena cava (IVC) diameter frames from a proximal IVC video used in deep learning algorithm training

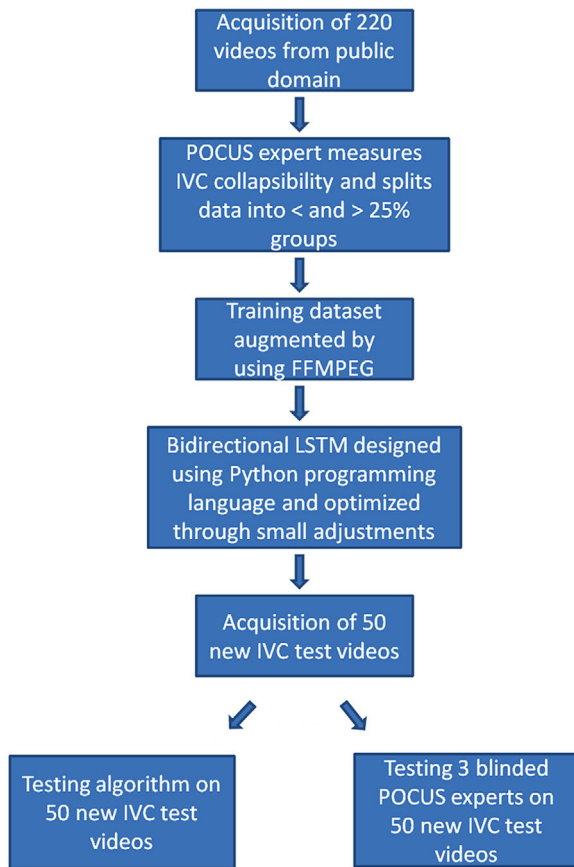


FIGURE 2 Diagram of a step-by-step approach for algorithm design, training, and testing

were used to guide algorithm training adjustments. After results were optimized, and no further adjustments resulted in performance improvements, the algorithm was tested on the 50 newly obtained IVC test videos, which were the same videos sent to POCUS experts for blinded review. No video augmentation was performed on the 50 test video set. Three POCUS experts (16, 12, and 11 years of POCUS experience, respectively) with fellowship training and extensive research, academic, and clinical scanning experience were asked to review the 50 IVC test videos in blinded fashion. Each POCUS expert received a drop-box link containing all 50 videos, randomly arranged and numbered. Experts were asked to quickly view each video and make a decision whether the IVC collapsed 25% or greater, looking at 3 cm distal to the diaphragm/right. The POCUS experts were asked to take a quick look (but were not timed) to better simulate real-life decisionmaking with POCUS where prolonged image review is prohibitive. Experts recorded their findings on an Excel spreadsheet for each IVC video. POCUS expert ratings were incorporated into 1 database and matched for each video along with the deep learning algorithm predictions on the same videos. Additionally, the 3 POCUS experts were asked to rate the difficulty of interpreting each IVC video to determine if it collapsed 25% or more on a 10-point Likert scale. Figure 2 summarizes the steps involved with algorithm construction, training, and testing.

2.6 | Statistical analysis

Fleiss' κ was calculated to compare level of agreement between the 3 POCUS experts and deep learning algorithm and POCUS experts. Fleiss' κ is a measure of inter-rater agreement used to determine the level of agreement between 2 or more raters. The single POCUS expert who rated degree of collapse for the 220 training IVC videos did not rate the 50 test videos, and there were no comparisons of the 3 POCUS experts and the deep learning to any criterion standard. Because there is no actual disease presence or absence, and the comparison of the deep learning prediction is compared to that of 3 POCUS experts, sensitivity, specificity, and likelihood ratios are not calculated because all the required components (true-positives, true-negatives, false-positives, and false-negatives) are not available. This is a common feature of imaging deep learning studies.

3 | RESULTS

All 3 POCUS experts were able to complete rating each of the 50 test IVC videos. Decisions regarding degree of collapse by each POCUS expert and the deep learning algorithm for each IVC video are listed in Table 1. Summary results relating to video classification into collapse and non-collapse groups are shown in Table 2. Asked to rate difficulty of interpretation of each video and quality, the experts reported no videos that could not be evaluated because of poor quality. Reported difficulty in IVC video interpretation by POCUS expert reviewer ranged from 1–7, 1–9, and 1–9, respectively. Median difficulty scores rating the difficulty POCUS reviewers had in determining if collapse was >25% or <25% for test IVC videos, and rating for IVC collapse and non-collapse subgroups are reported in Table 2. Percentage of IVC videos rated as showing 25% or greater collapse for the deep learning and 3 POCUS reviewers are shown in Table 2.

The LSTM deep learning algorithm was able to provide a prediction for each of the 50 videos. The algorithm took 9:24 min to train on the 220 augmented training videos. It took 30 seconds to review and make predictions on all 50 test IVC videos. Reviewers were not asked to time how long it took them to review and make a decision regarding collapse for the 50 test videos. There was very good agreement between the 3 POCUS experts with $\kappa = 0.65$ (95% CI = 0.49–0.81). Agreement between experts and algorithm was good with $\kappa = 0.45$ (95% CI = 0.33–0.56).

4 | DISCUSSION

Our results indicate that a deep learning neural network using LSTM and trained on public domain IVC ultrasound videos can achieve good agreement with POCUS experts in visually estimating if patient IVC collapsed 25% or greater. This cutoff is similar to other suggested ones and is supported by recent studies as being useful for determining if septic shock patients will be fluid responsive or not.^{6,7} Only a small number of studies have been published to date on either deep learning

TABLE 1 Three POCUS expert reviewer assessments as well as the DL prediction of the degree of IVC collapse for each test IVC video

IVC video number	POCUS 1	POCUS 2	POCUS 3	DL algorithm
IVC 1	Collapsing	Collapsing	Collapsing	Collapsing
IVC 2	Collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 3	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 4	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 5	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 6	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 7	Collapsing	Collapsing	Collapsing	Collapsing
IVC 8	Collapsing	Collapsing	Not collapsing	Not collapsing
IVC 9	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 10	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 11	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 12	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 13	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 14	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 15	Not collapsing	Collapsing	Not collapsing	Not collapsing
IVC 16	Not collapsing	Collapsing	Collapsing	Collapsing
IVC 17	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 18	Collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 19	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 20	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 21	Collapsing	Not collapsing	Not collapsing	Collapsing
IVC 22	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 23	Collapsing	Collapsing	Collapsing	Collapsing
IVC 24	Collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 25	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 26	Not collapsing	Not collapsing	Collapsing	Collapsing
IVC 27	Collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 28	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 29	Collapsing	Not Collapsing	Not Collapsing	Not collapsing
IVC 30	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 31	Collapsing	Not collapsing	Not collapsing	Collapsing
IVC 32	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 33	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 34	Collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 35	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 36	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 37	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 38	Collapsing	Collapsing	Collapsing	Collapsing
IVC 39	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 40	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 41	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 42	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 43	Collapsing	Collapsing	Collapsing	Collapsing
IVC 44	Collapsing	Not collapsing	Collapsing	Collapsing

(Continues)

TABLE 1 (Continued)

IVC video number	POCUS 1	POCUS 2	POCUS 3	DL algorithm
IVC 45	Collapsing	Collapsing	Collapsing	Not collapsing
IVC 46	Collapsing	Collapsing	Collapsing	Collapsing
IVC 47	Collapsing	Collapsing	Collapsing	Collapsing
IVC 48	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 49	Not collapsing	Not collapsing	Not collapsing	Not collapsing
IVC 50	Not collapsing	Not collapsing	Not collapsing	Not collapsing

DL, deep learning; IVC, inferior vena cava; POCUS, point-of-care ultrasound.

TABLE 2 DL prediction and 3 POCUS reviewer assessments of IVC collapsibility for the 50 test IVC videos

	POCUS reviewer 1	POCUS reviewer 2	POCUS reviewer 3	DL algorithm
Videos judged as collapsing $\geq 25\%$	30 (60%)	23 (46%)	23 (46%)	12 (24%)
Mean difficulty rating for all IVC videos	2.7 (95% CI = 2.06–3.42)	3.0 (95% CI = 2.42–3.5)	3.8 (95% CI = 2.93–4.67)	
Mean difficulty rating for IVC videos with collapse $\leq 25\%$	2.0 (95% CI = 1.34–2.58)	2.8 (95% CI = 2.13–3.47)	3.5 (95% CI = 2.39–4.65)	
Mean difficulty rating for IVC videos with collapse $< 25\%$	3.4 (95% CI = 2.27–4.53)	3.2 (95% CI = 2.23–4.17)	4.1 (95% CI = 2.72–5.41)	

DL, deep learning; IVC, inferior vena cava; POCUS, point-of-care ultrasound. The 10-point Likert scale ratings of difficulty interpreting IVC test videos are listed for each of the POCUS reviewers.

or other automated analysis of the IVC. The earliest ones predate modern deep learning techniques, which are only several years old. Our purpose in this study was not to compare the deep learning algorithm and 3 POCUS experts to any criterion standard, such as the original POCUS reviewer of the 220 training videos for the deep learning algorithm. Instead, our goal was to assess the internal consistency of the 3 POCUS expert reviewers and the deep learning algorithm.

IVC collapsibility has been explored in several published research studies on automatic interpretation and analysis. The first commercially available automated IVC analysis application by a POCUS ultrasound vendor was introduced nearly 4 years ago, but was not originally designed using artificial intelligence.¹⁸ The goal of any automation is to enable less experienced health care providers and those with less medical training to be able to assess volume status and potentially fluid responsiveness in sick patients who may be suffering from some type of shock. A group of researchers specifically focused on eliminating the time and labor intensive IVC measurement burden in clinical settings by creating and testing an automated process in 8 pigs.¹⁹ A total data set of 48 IVC evaluations was generated from the animals, and researchers were able to create a method to automatically identify and measure the IVC, offline and not in real-time. Although this is an important laboratory step forward, the authors describe a lengthy stepwise process required to manipulate the ultrasound cine loops with various filters and image size adjustment, limiting any real-

time clinical application potential. Despite the limited data, this automated method accurately identified 97.9% of pig IVCs. The majority of IVC measurements were within 15% of those made by 2 sonographers. A final limitation was that the methodology required DICOM data and took place offline, unless imbedded on an ultrasound device—that was not tested.¹⁹

An important collaboration between clinicians and engineering experts resulted in the creation of an automated technique for IVC diameter change assessment using a complex pyramidal structure that resulted in good agreement with manual measurements. The authors reported that when measurements between physicians and the algorithm differed, in 95% of these cases the difference was $< 10\%$.²⁰ The complex algorithm performed well when tested in 50 hemodialysis patients. However, it required multiple steps and significant engineering expertise to design and implement. Rather than undertake complex additional steps, which will inherently increase the difficulty of implementing such an algorithm on a wide range of POCUS machines, we sought to use the ability of deep learning algorithms to find novel associations to make predictions based on regression (referring to the type of deep learning methodology used as opposed to image segmentation and actual measurement). Therefore, no actual anatomic localization, border identification, or diameter measurements are required for prediction of outcome. Further, rather than requiring software to extract optimal individual images or

breaking ultrasound cine loops into individual images to analyze offline, we explored a real-time application that could analyze a cine loop immediately on an ultrasound machine or even run in real-time while a novice sonologists was scanning the patient's IVC.

Enabling providers new to POCUS, such as residents, untrained faculty, nurses, emergency medical technicians, and others to accurately assess the IVC through automation may significantly improve patient assessment, management, and access to care, and hasten interventions. Additionally, real-time feedback will further help novice sonologists improve their skills. A published study on nurses measuring the IVC diameter in the emergency department, following 3.5 hours of didactic and hands-on training, examined the correlation with a sonographer in both longitudinal and transverse IVC diameter measurements: $R = 0.68$ and 0.59 , respectively, in an all-volunteer model.²¹ Although this study proved that nurses can measure the IVC, correlation with a sonographer was suboptimal, and the training time required might have been better spent focusing on simply on attaining a mid-sagittal IVC view and holding it steady. As with many medical and non-medical tasks, deep learning-based automation should significantly decrease the amount of training required for nurses in this task. Additionally, evidence shows implemented artificial intelligence can significantly improve sonographer intra-observer reliability during cardiac measurements, suggesting artificial intelligence should accomplish this for novice ultrasound users also.¹⁰

Our efforts differ from prior designs and should offer easier implementation into clinical use when compared to the challenging process of programming an application to extract individual frames, performing multiple measurements and doing so offline. We modeled our approach on a POCUS expert's ability to rapidly view the proximal IVC and determine visually if 25% collapse occurs or not. This will allow the resultant deep learning application to be activated once the imaging window is obtained, hopefully yielding accurate results. Despite a small dataset for deep learning purposes, we were able to reach good agreement with 3 POCUS experts, suggesting that with a larger dataset correlating IVC behaviors on ultrasound and outcomes of fluid responsiveness testing, should be able to equal or surpass the results posted by expert POCUS. By embedding such a deep learning application into an ultrasound machine or simply having the application run on a real-time video feed from the system and displaying over the machines screen, a novice would only have to obtain an sagittal view of the proximal IVC and let the deep learning application do the rest, providing a nearly instant prediction regarding fluid responsiveness. The type of automation of IVC analysis we targeted would be of greater practical use in aiding novice POCUS providers than the current manual process. Rather than having to acquire a mid-sagittal image of the proximal IVC, then holding the transducer in the same anatomic position through several respiratory cycles, freezing the image and scrolling through the cine loop to identify the best IVC maximum and minimum diameters, and finally making a total of 4 measurements to assure measurement of the IVC diameter at the same anatomic point, the novice would instead be tasked only with obtaining the image and holding still over the anatomy.

Our study had a number of limitations including a very small dataset. Ideally, the dataset would have been in the thousands of videos and

would likely have resulted in much better correlation. Additionally, if the videos were in DICOM format, and we were able to obtain accurate distance and diameter measurements for training, the precision of the algorithm in differentiating between 25% collapse or greater versus less would be expected to grow significantly. Thus, the criterion standard used of a single POCUS expert to review and assess degree of IVC collapse itself introduces some amount of error. All of the videos were obtained from public domain sources, and although from actual patients, they varied in quality as did the equipment used to obtain them. It was impossible to obtain accurate patient states for all videos, and it is possible the 220 training and 50 test videos represented an atypical set of images. Data augmentation significantly increased the effective dataset size. However, for optimal results, only 1 type of ultrasound machine would have provided a very large dataset used to train the deep learning algorithm. This is not a realistic approach clinically, because many hospitals and even departments rely on a variety of ultrasound machines, therefore necessitating a robust algorithm trained on different ultrasound devices. Although tested on a separate set of ultrasound videos not previously seen by the artificial intelligence algorithm, we did not have other data from different medical centers nor were we able to test it prospectively in different settings. This would be an important future research topic, but this pilot study lays an important foundation for this approach.

In conclusion, a non-commercial LSTM deep learning algorithm in this study showed good agreement with 3 blinded POCUS expert reviewers in determining degree of IVC collapsibility in ultrasound videos of spontaneously breathing patients. This study indicates that an artificial intelligence algorithm has the potential to improve CIVC measurements.

AUTHOR CONTRIBUTIONS

MB and LNB conceived the study concept, gathered data, performed image classification, and wrote the article. MB and LNB created algorithmic approach and performed algorithm training and testing. SA, EAS, and YTL assisted in project creating, acted as blinded ultrasound reviewers, and helped write the Introduction and Discussion. MB takes the final responsibility of the article.

CONFLICTS OF INTEREST

MB consults with EchoNous Inc., Sonosim Inc. Ethos Medical and 410Medical. None of the companies had influence or contribution to this study and article or knowledge of its performance.

REFERENCES

1. Dipti A, Soucy Z, Surana A, Chandra S. Role of inferior vena cava diameter in assessment of volume status: a meta-analysis. *Am J Emerg Med.* 2012;30(8):1414-1419.e1.
2. Basu S, Sharron M, Herrera N, Mize M, Cohen J. Point-of-care ultrasound assessment of the inferior vena cava in mechanically ventilated critically ill children. *J Ultrasound Med.* 2020.
3. Orso D, Paoli I, Piani T, Cilenti FL, Cristiani L, Guglielmo N. Accuracy of ultrasonographic measurements of inferior vena cava to determine fluid responsiveness: a systematic review and meta-analysis. *J Intensive Care Med.* 2020;35(4):354-363.

4. Kusaba T, Yamaguchi K, Oda H. Echography of the inferior vena cava for estimating fluid removal from patients undergoing hemodialysis. *Nihon Jinzo Gakkai Shi*. 1996;38(3):119-123.
5. Bowra J, Uwagboe V, Goudie A, Reid C, Gillett M. Interrater agreement between expert and novice in measuring inferior vena cava diameter and collapsibility index. *Emerg Med Australas*. 2015;27(4):295-299.
6. Corl KA, George NR, Romanoff J, et al. Inferior vena cava collapsibility detects fluid responsiveness among spontaneously breathing critically-ill patients. *J Crit Care*. 2017;41:130-137.
7. Corl KA, Azab N, Nayeemuddin M, et al. Performance of a 25% inferior vena cava collapsibility in detecting fluid responsiveness when assessed by novice versus expert physician sonologists. *J Intensive Care Med*. 2019;885066619881123.
8. Fields JM, Lee PA, Jenq KY, Mark DG, Panebianco NL, Dean AJ. The interrater reliability of inferior vena cava ultrasound by bedside clinician sonographers in emergency department patients. *Acad Emerg Med*. 2011;18(1):98-101.
9. Akkaya A, Yesilaras M, Aksay E, Sever M, Atilla OD. The interrater reliability of ultrasound imaging of the inferior vena cava performed by emergency residents. *Am J Emerg Med*. 2013;31(10):1509-11.
10. Liao Z, Girgis H, Abdi A, et al. On modelling label uncertainty in deep neural networks: automatic estimation of intra-observer variability in 2D echocardiography quality assessment. *IEEE Trans Med Imaging*. 2019;39(6):1868-1883.
11. Singh R, Kalra MK, Nitiwarangkul C, et al. Deep learning in chest radiography: detection of findings and presence of change. *PLoS One*. 2018;13(10):1-12.
12. Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med*. 2018;15(11):1-19.
13. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392(10162):2388-2396.
14. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation*. 2018;138(16):1623-1635.
15. Asch FM, Poilvert N, Abraham T, et al. Automated Echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circ Cardiovasc Imaging*. 2019;12(9):1-9.
16. AI-Guided Ultrasound System Now Available in the U.S. News: Artificial Intelligence. 2020
17. Blaivas M, Blaivas L. Are all deep learning architectures alike for point-of-care ultrasound?: evidence from a cardiac image classification model suggests otherwise. *J Ultrasound Med*. 2019;39(6):1187-1194.
18. Venue Go R2 Technical Product Claims Document (DOC2199650).
19. Chen J, Li J, Ding X, Chang C, Wang X, Ta D. Automated identification and localization of the inferior vena cava using ultrasound: an Animal Study. *Ultrason Imaging*. 2018;40(4):232-244.
20. Belmont B, Kessler R, Theyyanni N, et al. Continuous inferior vena cava diameter tracking through an iterative kanade-lucas-tomasi-based algorithm. *Ultrasound Med Biol*. 2018;44(12):2793-2801.
21. De Lorenzo RA, Holbrook-Emmons VL. Ultrasound measurement of inferior vena cava diameters by emergency department nurses. *Adv Emerg Nurs J*. 2014;36(3):271-8.

AUTHOR BIOGRAPHY



Michael Blaivas, MD, MBA is an affiliate faculty at the University of South Carolina School of Medicine and practices clinically at St. Francis Hospital in Columbus, GA.

How to cite this article: Blaivas M, Adhikari S, Savitsky EA, Blaivas LN, Liu YT. Artificial intelligence versus expert: a comparison of rapid visual IVC collapsibility assessment between POCUS experts and a deep learning algorithm. *JACEP Open*. 2020;1:857-864. <https://doi.org/10.1002/emp2.12206>