Submitted: 12 May, 2021 Accepted: 23 June, 2021 Online Published: 14 July, 2021 DOI:10.22514/sv.2021.128

Open Access

Signa Vitae

ORIGINAL RESEARCH

Deep learning algorithm performance compared to experts in visual evaluation of interior vena cava collapse on ultrasound to determine intravenous fluid need in dehydration management

Michael Blaivas[1,*], Laura N Blaivas[2], James W Tsung[3]

[1]Department of Medicine, University of South Carolina School of Medicine, Department of Emergency Medicine, St. Francis Hospital, Columbus, GA 769209, USA
[2]Michigan State University, East Lancing, MI 48825, USA
[3]Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

***Correspondence**
mike@blaivas.org (Michael Blaivas)

Abstract

Objectives: To create a deep learning (DL) algorithm capable of analyzing real time ultrasound video of the inferior vena cava (IVC) for complete collapse in pediatric patients being evaluated for intravenous fluid (IVF) resuscitation.
Methods: Researchers employed a VGG-16 based DL architecture, running inside a Long Short Term Memory algorithm design, to analyze prospectively obtained ultrasound video from pediatric patients presenting with dehydration to a busy urban ED, obtained for a prior clinical study. All videos were de-identified and no patient information was available. A total of 184 patient IVC ultrasound videos were used in the study. All videos were previously reviewed and graded by two blinded POCUS experts (PedEM resident and PedEM attending with 20 years experience) and split into two categories, those showing complete (95 patients) and those with incomplete (89 patients) IVC collapse. Approximately 10% (9) patient videos were randomly removed from each original data groups to be used for algorithm testing after training was completed. A standard 80%/20% training and validation split was used on the remaining 166 patient videos for algorithm training. Training accuracy, losses and learning curves were tracked and various training parameters such as learning rates and batch sizes were optimized throughout training. As a final real world test, the DL algorithm was tasked with analyzing the 18 previously unseen, randomly selected IVC videos. Cohen's kappa was calculated for each of the blinded POCUS reviewers and DL algorithm.
Results: This DL algorithm completed analysis of each previously unseen real world test video and is the first such algorithm to analyze IVC collapse through visual estimation in real-time. The algorithm was able to deliver a collapse result prediction for all 18 test IVC videos and there were no failures. Algorithm agreement with PedEM POCUS attending was substantial with a Cohen's kappa of 0.78 (95% CI 0.49 to 1.0). Algorithm agreement with PedEM resident was substantial with Cohen's kappa of 0.66 (95% CI 0.31 to 1.0). The PEM resident and PEM POCUS attending also had substantial agreement, yielding a Cohen's kappa of 0.66 (95% CI 0.32 to 1.0).
Conclusions: This DL algorithm developed on prospectively acquired IVC video data from patients being studied for an IVF resuscitation study proved accurate at identifying when the IVC collapsed completely in real time. There was substantial agreement with POCUS reviewers of the same videos. Such an algorithm could allow novice clinicians to rapidly identify complete IVC collapse in children and the need for IVF administration. This could expand patient access to point of care technology by enabling novices with little training to use the diagnostic tool at bedside and decide if patients require intravenous fluid administration.

Keywords

Deep learning; Artificial intelligence; Long short term memory; Point-of-care ultrasound; Emergency medicine; Critical care; Inferior vena cava; Fluid responsiveness

This is an open access article under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

©2021 The Author(s). Published by MRE Press.

1. Introduction

Volume depletion secondary to various underlying causes is a frequent scenario in pediatric emergency patients [1]. In fact, worldwide, diarrhea claims tens of thousands of children's lives through dehydration [2]. However, determining whether a child has significant enough volume depletion to warrant intravenous fluid resuscitation or will simply recover with oral hydration can be surprisingly challenging for many providers [3]. Placing an intravenous line into a child can be difficult and traumatic for patient and parent alike, thus making unnecessarily placed lines even more significant in cases where oral hydration was safe to pursue and likely to work [4]. Signs which may point to volume depletion in adults can be absent in children until they are critically ill [5]. Scores and decision rules, designed to differentiate if young children are significantly dehydrated or not, are not uniformly used and have questionable clinical utility [6].

Ideally, all pediatric patients at risk for significant volume depletion could be noninvasively screened and the degree of hypovolemia identified or at least the need for intravenous fluids ruled out. Point-of-care ultrasound (POCUS) offers such a tool and the most common POCUS approach is interrogation of the inferior vena cava (IVC). Evaluation of the IVC has been shown to be a reliable indicator of volume status and results in good interrater reliability [7]. The largest body of IVC and volume status or fluid responsiveness literature has come from adult patients. In adults, measuring collapsibility of the inferior vena cava (cIVC) has been shown to predict fluid responsiveness fairly well, when performed in non-ventilated patients by expert sonologists (AUROC = 0.82) [8]. However, work by another group of authors raised concerns that the test characteristics were not so favorable when novice sonologists performed the POCUS examination (AUROC = 0.69) [9]. Other researchers have also reported, poor to moderate interrater agreement in cIVC measurement among non-expert sonologists [10–12]. Training clinicians with no ultrasound experience, to a level at which they attain competence with POCUS use and accurately measure cIVC to guide IV fluid resuscitation, is emerging as the most significant hurdle to broad adoption of this approach.

Prior work suggests that euvolemic children will not have complete IVC collapse on inspiration, thus on ultrasound the anterior and posterior walls never touch in a sagittal plane throughout the proximal IVC [13]. Such patients can be treated without intravenous fluid infusion and be discharged home. However, significant collapse is more frequently seen in children with dehydration. One practical limitation to POCUS evaluation of the IVC is the skill required to not only obtain the correct image but then rapidly interpret IVC behavior in real time at the patient's bedside. This may be made more difficult in a busy environment or with a sick, poorly cooperative child. Published studies have shown that a good cutoff point for determining if a child requires intravenous fluid administration is the presence of complete IVC collapse on ultrasound —suggesting IVF is required [13]. While, on the surface, appearing simple by description, the procedure can be difficult for novice ultrasound users. A tool to automatically identify when the IVC collapses completely could prove useful in broader clinical practice. Adapting deep learning (DL) as a solution to overcome the lack of adequate training, thereby enabling widespread use of POCUS targeted resuscitation more feasible, has been proposed [14].

In this study, researchers explored the development of a Deep Learning (DL)/Artificial Intelligence (AI) algorithm which would identify occurrence of complete IVC collapse. Accurate and rapid identification of either the presence or lack of complete IVC collapse would allow novice users to identify children likely to be go home without bounce back for dehydration. A DL algorithm was created using ultrasound videos of the IVC in pediatric patients being evaluated for volume status, to identify those with complete IVC collapse among spontaneously breathing patients. Researchers then compared the performance of the DL algorithm against 2 POCUS experts on new (not previously part of the DL training dataset) IVC videos.

2. Methods

2.1 Study design

Researchers used de-identified ultrasound video data from a prior prospective observational study of patients ranging up to 21 years of age, who presented to the ED [15]. Videos of IVC scans were obtained for patients with a history of diarrhea with suspected dehydration and or a history of emesis being treated with ondanseton. Patients who were unstable were not included in this dataset. Patient were scanned after being triaged, but prior to fluid resuscitation or rehydration. Researchers performed an IVC ultrasound examination which included obtaining and recording a sagittal view of the proximal IVC during several respiratory cycles. This study was approved by the Icahn School of Medicine at Mount Sinai institutional review board, reference number 15-1599.

Providers used a Sonosite M-Turbo ultrasound system with a phased array transducer (P21 -5 to 1 MHz). The researchers used a cardiac preset for imaging the IVC and saved 6 second retrospective cine loops, Fig. 1. Complete IVC collapse was defined as opposing walls of the IVC coming into contact with each other at any point, Fig. 2. A total of 5 researchers enrolled patients and scanned IVCs, 4 pediatric emergency medicine (PEM) attendings and 1 PEM resident. The researchers scanned with the transducer in the subxiphoid location using the liver as an acoustic window. A sagittal plane over the mid proximal IVC was attained, visualizing the IVC entry into the right atrium, Fig. 1. Researchers focused their attention just caudal to the junction of the hepatic veins with the IVC.

2.2 Study data

A total of 166 individual patient videos of the IVC were available for algorithm training. A standard 80%/20% training and validation split was applied to the 166 patient videos for algorithm training. Video data was in MP4 format. Each video was reviewed by two PEM POCUS providers and rated as completely collapsing or not. Complete collapse was defined as the anterior and posterior walls of the proximal IVC touching at any point during inspiration. An additional new 18 videos, data

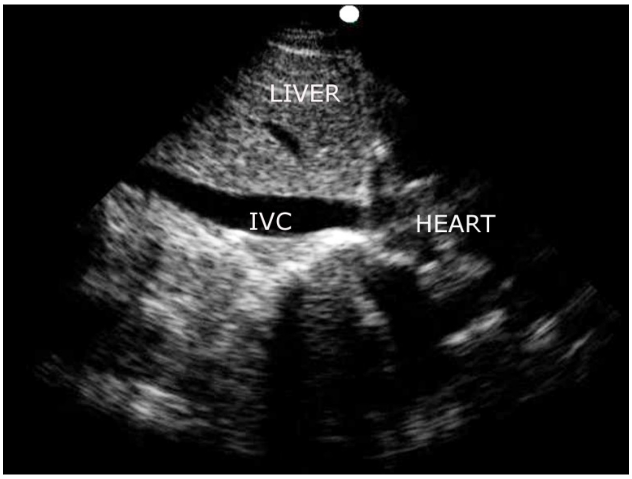

FIGURE 1. Shows sagittal proximal IVC (inferior vena cava) segment appearance on ultrasound at maximum expiration in one of the study patients.

not seen previously by the algorithm and not from the same training set, were used for algorithm testing as real world data, Fig. 3.

2.3 Algorithm design

Researchers used the Anaconda package manager with Python programming language version 3.72, to manage packages and facilitate scripting. Based on prior experience, a decision was made to utilize a publicly available Keras-based (a python deep learning library or framework) VGG-16 bidirectional LSTM DL algorithm. The VGG-16 model can be accessed from several public sources, including github.com (an online scripting repository). An early version of a convolutional neural network (CNN), VGG-16 uses 16 layers and has been shown to be superior for ultrasound DL applications when compared to modern and more complex CNNs [9]. The LSTM architecture allows analysis of individual ultrasound frames by the VGG-16 CNN, while it tracks temporal changes and relationships in real-time video to create an algorithm for ultrasound video analysis, Fig. 4. An LSTM such as this one has additional layers in the network algorithm architecture which tracks temporal changes between images (changes in the ultrasound image in a cine loop such as a beating heart). In contrast to standard LSTM networks, which are unidirectional, the bidirectional aspect allows temporal related information to flow in both forward and reverse directions through the algorithm. This feature makes the architecture more sensitive and specific for detecting any changes from frame to frame within a video, thereby better identifying action. It further improves a network's understanding of the context the motion occurs in. Investigators used weights trained on the UCF-101 Action Recognition Data Set (University of Central Florida) for the VGG-16 bidirectional LSTM. Weights are learnable parameters in neural networks and account for a CNN's ability to interpret images.

This appears to be the first instance of LSTM DL architecture being applied to pediatric IVC measurement. The majority of prior LSTM work has focused on ECG and EEG analysis as well as computed tomography image analysis [16–18]. These imaging types pose different challenges to analysis than due to limited number of frames requiring analysis and change occurring over a short period of time. Assessing pediatric IVC behavior in acute settings results in considerable movement due to rapid breathing, heart rate and in younger patients poor cooperation. LSTM has been previously applied to adult IVC assessment using similar technique, but adults have differing physiology and larger anatomic size than children [19, 20]. LSTM architecture has been used for needle tip analysis and possible ultrasound guidance but only on non-human tissue in a laboratory setting [21]. The closest other LSTM work with ultrasound targeted lung ultrasound analysis in adults but utilized a mono-directional LSTM network with DenseNet 201 embedded in the architecture. However, the VGG 16 utilized in the current approach may be more efficacious when analyzing ultrasound images [15, 22].

The bidirectional LSTM algorithm was trained by incrementally manipulating learning rates, optimizers, batch size and learning rates (which are all adjustable settings which affect the performance of a CNNs during training) during training for optimal accuracies and training times. Simultaneously researchers worked to avoid exploding gradients (dramatic changes in learning parameters of the CNN during training) which can result in training failure. The total epochs used, defined as one round of training through all of the data, was adjusted as needed to optimize results while avoiding overfitting. Analyzing 90 frame runs proved to be the most efficacious approach.

2.4 Algorithm validation and testing

The bidirectional LSTM algorithm was written to perform cross validation during each epoch automatically. Learning losses, training losses and cross-validation accuracy, were used to guide algorithm training adjustments. Best performance was obtained when training for 63 epochs (epoch = single training pass through all of the data), using a Stochastic gradient decent optimizer a learning rate of 0.001 and batch size of 10 videos. After results were optimized and no further adjustments improved performance, the algorithm was tested on the 18 completely new, prospectively obtained videos from different patients. A testing script used the newly generated training weights on the 18 new patient IVC ultrasound videos through the VGG-16 bidirectional LSTM to predict IVC collapse. This step tested the algorithms' performance in real life by applying it to newly, prospectively obtained data, not previously used to train the DL algorithm. Researchers then compared these results to 2 POCUS experts who reviewed the same videos. The expert physician sonologists were an attending PEM attending physician and a PEM resident at the end of their fellowship with extensive ultrasound experience, both had performed over 250 IVC ultrasound examinations.

2.5 Statistical analysis

Cohen's kappa (κ) with 95% confidence intervals (CIs) was calculated between the 2-blinded reviewers, and between the reviewers and the DL algorithm. Statistical analyses were performed using IBM SPSS Statistics 25 (IBM, Armonk, NY, USA).

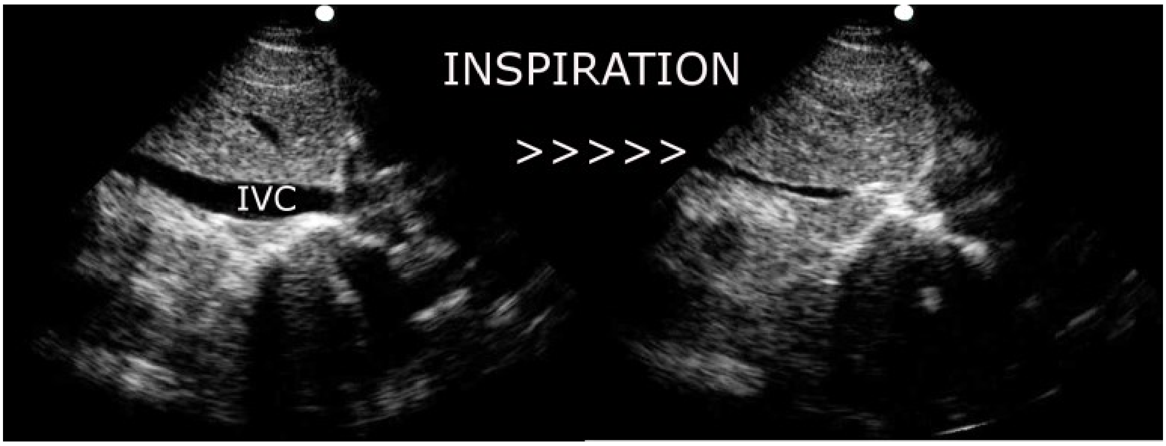

FIGURE 2. Shows a patient whose IVC collapses completely upon inspiration from left to right. Complete IVC collapse was defined as opposing walls of the IVC coming into contact with each other at any point.

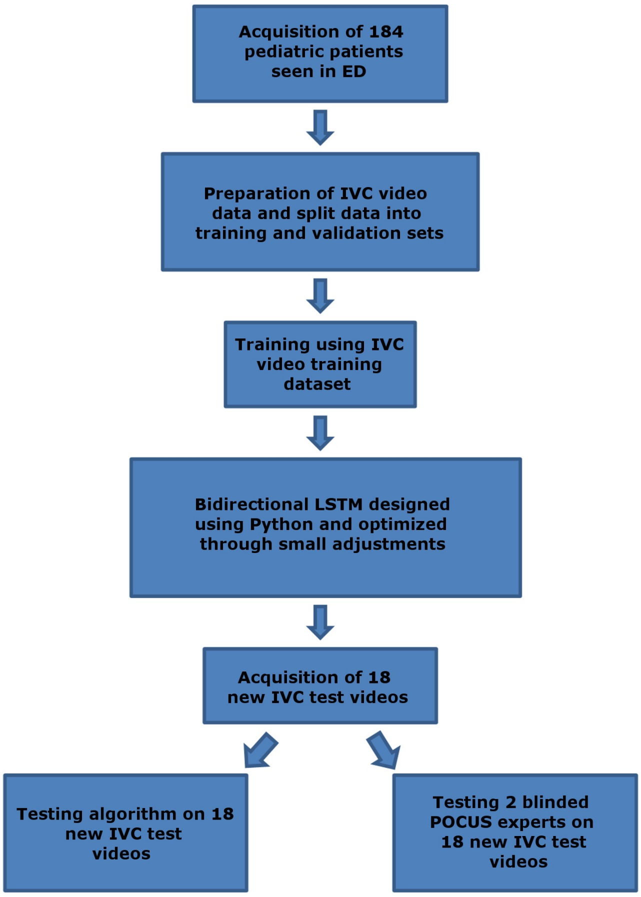

FIGURE 3. The study process flow in graphical form.

3. Results

3.1 Outcomes

Table 1 provides reviewers' interpretation of collapse and DL algorithm's prediction of collapse completeness for all 18 videos. The DL algorithm and both reviewers were able to complete review and categorization of each test video. The POCUS reviewers felt all videos were adequate for collapse assessment. None of the patients who had partial IVC collapse required IVF resuscitation or admission solely for dehydration. All patients with incomplete IVC collapse (Fig. 5) were able to be discharged from the ED and did not return to the same medical system, based on chart review. This DL algorithm was able to complete analysis of each previously unseen real world test video. Algorithm agreement with the PEM POCUS attending was substantial with a Cohen's kappa of 0.78 (95% CI 0.49 to 1.0). Algorithm agreement with PEM resident was substantial with Cohen's kappa of 0.66 (95% CI 0.31 to 1.0). The PEM resident and PEM POCUS attending also had substantial agreement, yielding a Cohen's kappa of 0.67 (95% CI 0.32 to 1.0).

The DL algorithm disagreed with the PEM POCUS attending in two patients. Both disagreements resulted from a prediction by the DL algorithm of complete IVC collapse, while the PEM POCUS attending felt the IVC did not collapse completely. The DL algorithm and PEM resident disagreed in three instances with the algorithm predicting complete IVC collapse twice while the resident rated the IVC as not collapsing completely. In the third disagreement the DL algorithm predicted incomplete collapse while the resident judged the IVC to collapse completely. In two out of the three disagreements between the DL algorithm and resident, the POCUS attending agreed with the algorithm. A review of the videos which caused disagreement did not reveal significant challenges with image artifact or shadowing involving the IVC when compared to the remaining test videos.

4. Discussion

In this study, researchers developed a novel AI algorithm for IVC ultrasound interpretation in pediatric emergency patients being evaluated for dehydration. Considerable prior work has appeared in the medical literature including comparison of the IVC diameter in maximum and minimum and its relationship with the aorta diameter [23]. This approach is quite laborious, requiring measurement of the aorta and IVC maximum and minimum diameters to establish IVC collapsibility and is rarely used in clinical practice, making it essentially obsolete [13]. However, even the process of IVC measurement for maximum and minimum diameters, to establish IVC collapsibility, takes time and requires additional ultrasound experience that some

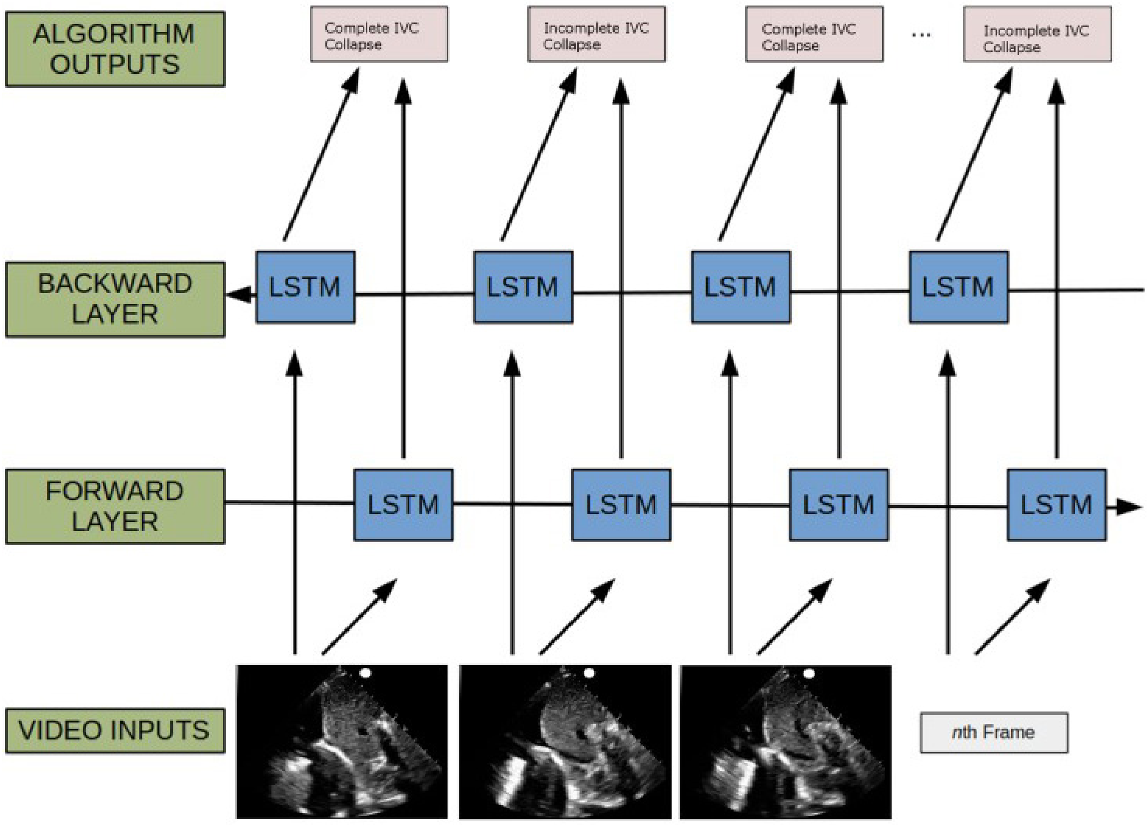

FIGURE 4. Shows the LSTM architecture developed and employed in the study.

TABLE 1. Comparison table of Deep Learning algorithm, PEM resident and PEM POCUS attending assessment of IVC collapse.

DL algorithm prediction	PEM resident	PEM POCUS attending
Incomplete collapse	Complete collapse	Incomplete collapse
Complete collapse	Complete collapse	Complete collapse
Incomplete collapse	Incomplete collapse	Incomplete collapse
Complete collapse	Complete collapse	Complete collapse
Incomplete collapse	Incomplete collapse	Incomplete collapse
Complete collapse	Complete collapse	Complete collapse
Incomplete collapse	Incomplete collapse	Incomplete collapse
Complete collapse	Complete collapse	Complete collapse
Complete collapse	Complete collapse	Complete collapse
Incomplete collapse	Incomplete collapse	Incomplete collapse
Complete collapse	Incomplete collapse	Incomplete collapse
Incomplete collapse	Incomplete collapse	Incomplete collapse
Complete collapse	Incomplete collapse	Complete collapse
Complete collapse	Complete collapse	Complete collapse
Complete collapse	Complete collapse	Complete collapse
Complete collapse	Complete collapse	Complete collapse
Incomplete collapse	Incomplete collapse	Incomplete collapse
Complete collapse	Complete collapse	Incomplete collapse

novices may not possess [24]. Prior work has also established that operators are able to visually categorize the degree of IVC collapse into discrete categories showing substantial interobserver agreement [25]. Additionally, investigators have previously shown that IVC collapse duration, when assessed by blinded reviewers, correlated well with PO versus IV re-

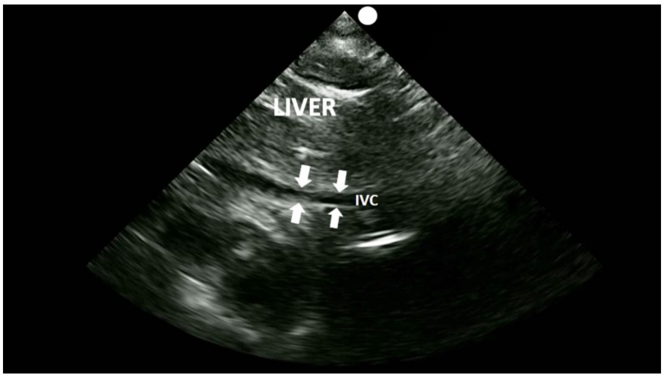

FIGURE 5. Shows a patient whose IVC collapsed incompletely upon inspiration.

hydration in pediatric patients [13]. Further, results indicated that lack of IVC collapsibility correlated with successful PO rehydration and discharge from the ED without patient bounce back [13].

Recently, attention has been focused on the potential utility of AI in POCUS applications and the promise for decreased cognitive burden in obtaining focused findings from ultrasound examinations by novices. More complex applications have also been explored and some are available for advanced ultrasound imaging to improve workflow and decrease examination times [26]. Creating a DL application which would allow clinicians with less ultrasound training to employ focused ultrasound in an equivalent way to POCUS experts would greatly increase patient access to a useful, non-invasive, bedside diagnostic and monitoring tool. While expert POCUS users are be able to quickly estimate degree of IVC collapse in a child on ultrasound, novices and less trained POCUS users may struggle to simply obtain and keep a mid-sagittal view of the proximal IVC. An algorithm which automatically "visually" interprets the degree of IVC collapse, without the novice operator having to take more steps, is directly in line with the promise of AI impact in clinical practice, especially imaging interpretation.

Most prior work with IVC assessment automation in ultrasound focused on animal models, created complex offline interpretation approaches with multiple steps required or compared AI results with carefully obtained measurements by ultrasound technologists [27]. Chen *et al*. [27], attempted to automate IVC collapse analysis and constructed a DL algorithm capable of identifying and measuring the IVC diameter in a porcine model. The DL algorithm was trained on 48 data sets of IVC images which included both dynamic and static images. Investigators then compared the DL algorithms IVC diameter measurements to those of expert sonographers. The researchers reported the algorithm successfully identified the IVC 98% of the time and most IVC diameter measurements made were within 15% of an expert sonologist's read. However, the researchers utilized a stepwise process that included significant captured video clip pre-processing, filters and image size modifications. Additionally, investigators relied on Color Doppler use to identify areas of interest based on blood flow presence alone, a technique typically unneeded in POCUS practice. Although promising, the multiple cumbersome steps in the described approach is not suitable for actual clinical situations. A more promising study was published by, Belmont et al. and used Kana-de-Lucas-Tomasi (KTL) feature tracking and pyramidal segmentation to build an artificial intelligence algorithm which was capable of identifying and measuring cIVC [28]. The somewhat complex algorithm analyzed 57 cine loops from 47 non-ventilated dialysis patients and showed good agreement with POCUS expert manual measurements. More than 95% of the DL measurements were within $<$10% compared to expert reads). Still, the algorithm's clinical application would likely face challenges in this format due to a requirement for clear and consistent IVC edge depiction in order to enable tracking. A requirement which can be difficult to maintain in the clinical setting.

A more practical approach is required, one that would focus on bedside application and tailored to enabling rapid interpretation in real time when used by ultrasound novices. Despite mental imagery conjured up by the term Deep Learning, if properly structured, DL can be implemented as a simple to use application and obviate the need for multiple offline steps. Deep Learning is the cornerstone of medical imaging analysis with artificial intelligence. The vast majority of work, both research and commercial, has focused on high dollar imaging modalities such as MRI and CT [29]. Plane x-rays have also generated considerable work, but ultrasound in general, lags far behind. Even less attention has been focused on POCUS related applications.

While seemingly magical in its capability to the uninitiated, DL is simply a massive amount of computations performed on ever larger portions of an image and correlating prediction outcomes with ground truth results, which indicate if the image is a cat or dog or in the case of medical images, gallbladder or heart, among others. Currently, there are at least two commercial products available, on ultrasound machines marketed in the USA, which have automatic IVC evaluation delivering a collapsibility index [20]. One of these uses a Deep Learning algorithm, which the other one is older and hard coded. Unfortunately, these algorithms are available on a small number of the ultrasound machines sold to POCUS providers and will not work with other ultrasound machine types, if they were transferable. Both of these automatic software applications depend on identifying the IVC and then tasking the software with marking IVC borders and then creating measurements from these, using M-mode tracings, followed by simple calculations. Calculation for both depends on excellent image quality, which may be very challenging in some clinical settings and is less likely to occur when ultrasound examinations are being performed by novices. The algorithm employed in this study uses a different property of Deep Learning networks, one that is similar to what experts do when they "eye-ball" an ultrasound assessment such as ejection fraction for the left ventricle. This is often referred to as the regression approach to DL algorithm function. Thus, the highest quality images with clear edges are not important for function to the same degree as current commercial versions of auto IVC evaluation tools. Additionally, the algorithm allows novices to focus on locating the proximal IVC and holding the transducer aligned over it. Researchers were able to achieve these results through a combination of a previously studied optimized CNN choice for ultrasound analysis and a fine tuned Bidirectional LSTM architecture which allowed for analysis of actual 6 second video clips of the IVC, the same as those used in clinical patient management. Therefore, this is the closest step to actually implementing and testing such an approach in an actual clinical setting.

The clinical impact of a DL algorithm that visually estimates IVC collapse is potentially significant. With fewer computational steps required there is increased efficiency and the possibility for a DL algorithm to run on simpler and smaller computer chips. For instance, a common approach for an algorithm such as this would be to take a live video feed from the ultrasound device and have the DL application running on a separate laptop or tablet in real time to assess the degree of IVC collapse. This approach implies that individual medical centers could design their own algorithms, tailored to the ultrasound devices available in their department and even fine-tuned to their patient population, which can change around the world. When a child presents to an emergency department, regardless of where in the world it is located, a novice ultrasound provider, nurse or tech, can assess volume status easily as long as they can obtain a view of the proximal IVC.

This study has multiple limitations. First the training sample size is considered quite small for DL algorithm design. A much larger dataset, one that is over 1000 individual patients would make for a more robust algorithm. Second, the test dataset used, (n = 18 patients) is also relatively small, but importantly allowed for a prospective, real life evaluation of algorithm performance, something typically absent in AI and imaging studies [30]. Implementing an executable application version of this algorithm was not performed by us, due to funding limitations and IRB permission, but is rudimentary and could be accomplished with a separate laptop that would take a video feed from an ultrasound device and in real time analyze and overlay results onto the ultrasound image. This application has not yet been tested in daily ED practice with various providers, a step for another future study, ideally once funding was available.

5. Conclusions

Researchers demonstrated that an DL algorithm trained on a small number of patient ultrasound cine loops can have substantial agreement with two blinded expert POCUS reviewers in judging complete versus incomplete IVC collapse in pediatric patient. Lack of complete IVC collapse is inversely correlated with the need for admission and IV fluid hydration, making this a useful practical tool in volume status assessment. Further, these results suggest that DL visual estimation of IVC collapse can be accurate and could be implemented by just one emergency department for their own use, even when commercial products are unavailable or unaffordable. This could expand patient access to point of care technology by enabling novices with little training to use the diagnostic tool at bedside.

AUTHOR CONTRIBUTIONS

MB, LB and JT were involved in research study design. MB and LB performed algorithm creation, training, testing and final predictions. JT, MB and LB analyzed the data and wrote the manuscript. All authors contributed to the editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Icahn School of Medicine at Mount Sinai institutional review board, reference number 15-1599.

ACKNOWLEDGMENT

Researchers would like to thank Xianshun Chen (chen0040) for posting example LSTM structure code on github: `https://github.com/chen0040/keras-video-classifier.git`.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

Michael Blaivas has consulted for the following commercial entities within the last 12 months: Ava AG, Anavasi Diagnostics, Sonosim Inc, EchoNous Inc, Ethos Medical. The remaining authors declare that there is no conflict of interest regarding the publication of this article.

DATA AVAILABILITY

The data used to support the findings of this study are available from the corresponding author upon request.

REFERENCES

[1] Alpern ER, Stanley RM, Gorelick MH, Donaldson A, Knight S, Teach SJ, *et al*. Pediatric Emergency Care Applied Research Network. Epidemiology of a pediatric emergency medicine research network: the PECARN Core Data Project. Pediatric Emergency Care. 2006; 22: 689–699.

[2] Modi P, Glavis-Bloom J, Nasrin S, Guy A, Chowa EP, Dvor N, *et al*. Accuracy of Inferior Vena Cava Ultrasound for Predicting Dehydration in Children with Acute Diarrhea in Resource-Limited Settings. PLoS ONE. 2016; 11: e0146859.

[3] Gorelick MH, Shaw KN, Murphy KO. Validity and reliability of clinical signs in the diagnosis of dehydration in children. Pediatrics. 1997; 99: E6.

[4] Yen K, Riegert A, Gorelick MH. Derivation of the DIVA score: a clinical prediction rule for the identification of children with difficult intravenous access. Pediatric Emergency Care. 2008; 24: 143–147.

[5] Goldman RD, Friedman JN, Parkin PC. Validation of the clinical dehydration scale for children with acute gastroenteritis. Pediatrics. 2008; 122: 545–549.

[6] Kinlin LM, Freedman SB. Evaluation of a Clinical Dehydration Scale

in Children Requiring Intravenous Rehydration. Pediatrics. 2012; 129: e1211–e1219.

[7] Pershad J, Myers S, Plouman C, Rosson C, Elam K, Wan J, *et al*. Bedside limited echocardiography by the emergency physician is accurate during evaluation of the critically ill patient. Pediatrics. 2005; 114: e667–e671.

[8] 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. Journal of Intensive Care Medicine. 2020; 35: 354–363.

[9] Corl KA, Azab N, Nayeemuddin M, Schick A, Lopardo T, Zeba F, *et al*. Performance of a 25

[10] Bussmann BM, Sharma S, Mcgregor D, Hulme W, Harris T. Observational study in healthy volunteers to define interobserver reliability of ultrasound haemodynamic monitoring techniques performed by trainee doctors. European Journal of Emergency Medicine. 2019; 26: 217–223.

[11] 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. Emergency Medicine Australasia. 2015; 27: 295–299.

[12] 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. The American Journal of Emergency Medicine. 2013; 31: 1509–1511.

[13] Zhou AZ, Green RS, Haines EJ, Vazquez MN, Tay ET, Tsung JW. Interobserver Agreement of Inferior Vena Cava Ultrasound Collapse Duration and Correlated Outcomes in Children with Dehydration. Pediatric Emergency Care. 2020. (in press)

[14] Shokoohi H, LeSaux MA, Roohani YH, Liteplo A, Huang C, Blaivas M. Enhanced Point-of-Care Ultrasound Applications by Integrating Automated Feature-Learning Systems Using Deep Learning. Journal of Ultrasound in Medicine. 2019; 38: 1887–1897.

[15] Blaivas M, Blaivas L. Are all Deep Learning Architectures Alike for Point-of-Care Ultrasound?: Evidence from a Cardiac Image Classification Model Suggests otherwise. Journal of Ultrasound in Medicine. 2020; 39: 1187–1194.

[16] Çınar A, Tuncer SA. Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks. Computer Methods in Biomechanics and Biomedical Engineering. 2021; 24: 203–214.

[17] Lai CQ, Ibrahim H, Abd Hamid AI, Abdullah JM. Classification of Non-Severe Traumatic Brain Injury from Resting-State EEG Signal Using LSTM Network with ECOC-SVM. Sensors. 2020; 20: 5234.

[18] Ko H, Chung H, Lee H, Lee J. Feasible Study on Intracranial Hemorrhage Detection and Classification using a CNN-LSTM Network. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2020; 2020: 1290–1293.

[19] Blaivas M, Blaivas L, Philips G, Merchant R, Levy M, Abbasi A, *et al*. Development of a Deep Learning Network to Classify Inferior Vena Cava Collapse to Predict Fluid Responsiveness. Journal of Ultrasound in Medicine. 2020. (in press)

[20] Blaivas M, Adhikari S, Savitsky EA, Blaivas LN, Liu YT. Artificial intelligence versus expert: a comparison of rapid visual inferior vena cava collapsibility assessment between POCUS experts and a deep learning algorithm. Journal of the American College of Emergency Physicians Open. 2020; 1: 857–864.

[21] Mwikirize C, Kimbowa AB, Imanirakiza S, Katumba A, Nosher JL, Hacihaliloglu I. Time-aware deep neural networks for needle tip localization in 2D ultrasound. International Journal of Computer Assisted Radiology and Surgery. 2021; 16: 819–827.

[22] Dastider AG, Sadik F, Fattah SA. An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound. Computers in Biology and Medicine. 2021; 132: 104296.

[23] Chen L, Kim Y, Santucci KA. Use of ultrasound measurement of the inferior vena cava diameter as an objective tool in the assessment of children with clinical dehydration. Academic Emergency Medicine. 2007; 14: 841–845.

[24] Kornblith AE, van Schaik S, Reynolds T. Useful but not used: pediatric critical care physician views on bedside ultrasound. Pediatric Emergency Care. 2015; 31: 186–189.

[25] Sawe HR, Haeffele C, Mfinanga JA, Mwafongo VG, Reynolds TA. Predicting Fluid Responsiveness Using Bedside Ultrasound Measurements of the Inferior Vena Cava and Physician Gestalt in the Emergency Department of an Urban Public Hospital in Sub-Saharan Africa. PLoS ONE. 2016; 11: e0162772.

[26] Mesin L, Pasquero P, Albani S, Porta M, Roatta S. Semi-automated tracking and continuous monitoring of inferior vena cava diameter in simulated and experimental ultrasound imaging. Ultrasound in Medicine & Biology. 2015; 41: 845–857.

[27] 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. Ultrasonic Imaging. 2018; 40: 232–244.

[28] Belmont B, Kessler R, Theyyunni N, Fung C, Huang R, Cover M, *et al*. Continuous Inferior Vena Cava Diameter Tracking through an Iterative Kanade–Lucas–Tomasi-Based Algorithm. Ultrasound in Medicine & Biology. 2018; 44: 2793–2801.

[29] Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, *et al*. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet. 2018; 392: 2388–2396.

[30] Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. Korean Journal of Radiology. 2019; 20: 405–410.

How to cite this article: Michael Blaivas, Laura N Blaivas, James W Tsung. Deep learning algorithm performance compared to experts in visual evaluation of interior vena cava collapse on ultrasound to determine intravenous fluid need in dehydration management. Signa Vitae. 2021. doi:10.22514/sv.2021.128.