

UCSF

UC San Francisco Previously Published Works

Title

Development and Validation of a Deep Learning Strategy for Automated View Classification of Pediatric Focused Assessment With Sonography for Trauma

Permalink

<https://escholarship.org/uc/item/0rt5469j>

Journal

Journal of Ultrasound in Medicine, 41(8)

ISSN

0278-4297

Authors

Kornblith, Aaron E
Addo, Newton
Dong, Ruolei
[et al.](#)

Publication Date

2022-08-01

DOI

10.1002/jum.15868

Peer reviewed



HHS Public Access

Author manuscript

J Ultrasound Med. Author manuscript; available in PMC 2022 August 01.

Published in final edited form as:

J Ultrasound Med. 2022 August ; 41(8): 1915–1924. doi:10.1002/jum.15868.

Development and Validation of a Deep Learning Strategy for Automated View Classification of Pediatric Focused Assessment With Sonography for Trauma

Aaron E. Kornblith, MD,

Department of Emergency Medicine, University of California, San Francisco, CA, USA

Department of Pediatrics, University of California, San Francisco, CA, USA

Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA

Newton Addo, BSc,

Department of Emergency Medicine, University of California, San Francisco, CA, USA

Department of Medicine, Division of Cardiology, University of California, San Francisco, CA, USA

Ruolei Dong, MTM,

Department of Bioengineering, University of California, Berkeley, CA, USA

Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA

Robert Rogers, PhD,

Center for Digital Health Innovation, University of California, San Francisco, CA, USA

Jacqueline Grupp-Phelan, MD, MPH,

Department of Emergency Medicine, University of California, San Francisco, CA, USA

Department of Pediatrics, University of California, San Francisco, CA, USA

Atul Butte, MD, PhD,

Department of Pediatrics, University of California, San Francisco, CA, USA

Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA

Pavan Gupta, BSc,

Center for Digital Health Innovation, University of California, San Francisco, CA, USA

Rachael A. Callcut, MD, MSPH,

Center for Digital Health Innovation, University of California, San Francisco, CA, USA

Department of Surgery, University of California, Davis, CA, USA

Rima Arnaout, MD

Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA

Address correspondence to Aaron E Kornblith, MD, 550 16th Street, Box 0632, San Francisco, CA 94143, USA.
aaron.kornblith@ucsf.edu.

Rachael A. Callcut and Rima Arnaout contributed equally as senior authors.

The authors declare no conflicts of interest.

Department of Medicine, Division of Cardiology, University of California, San Francisco, CA, USA

Abstract

Objective—Pediatric focused assessment with sonography for trauma (FAST) is a sequence of ultrasound views rapidly performed by clinicians to diagnose hemorrhage. A technical limitation of FAST is the lack of expertise to consistently acquire all required views. We sought to develop an accurate deep learning view classifier using a large heterogeneous dataset of clinician-performed pediatric FAST.

Methods—We developed and conducted a retrospective cohort analysis of a deep learning view classifier on real-world FAST studies performed on injured children less than 18 years old in two pediatric emergency departments by 30 different clinicians. FAST was randomly distributed to training, validation, and test datasets, 70:20:10; each child was represented in only one dataset. The primary outcome was view classifier accuracy for video clips and still frames.

Results—There were 699 FAST studies, representing 4925 video clips and 1,062,612 still frames, performed by 30 different clinicians. The overall classification accuracy was 97.8% (95% confidence interval [CI]: 96.0–99.0) for video clips and 93.4% (95% CI: 93.3–93.6) for still frames. Per view still frames were classified with an accuracy: 96.0% (95% CI: 95.9–96.1) cardiac, 99.8% (95% CI: 99.8–99.8) pleural, 95.2% (95% CI: 95.0–95.3) abdominal upper quadrants, and 95.9% (95% CI: 95.8–96.0) suprapubic.

Conclusion—A deep learning classifier can accurately predict pediatric FAST views. Accurate view classification is important for quality assurance and feasibility of a multi-stage deep learning FAST model to enhance the evaluation of injured children.

Keywords

abdominal injuries/diagnostic imaging; ultrasonography; pediatric trauma; machine learning; deep learning

Focused assessment with sonography for trauma (FAST) is a diagnostic point-of-care ultrasound study, in which a sequence of ultrasonographic views performed by the bedside treating clinician rapidly identifies free fluid, interpreted as hemorrhage in the setting of injury. FAST decreases computed tomography usage and improves clinical outcomes in adults with blunt injury,¹ but has shown inconsistent diagnostic utility in children.^{2,3} This inconsistency of FAST in children arises from varying image acquisition and interpretation of all the possible views.^{4–6} Specifically, non-interpretable FAST views are a major source of technical error.⁷ Technical expertise improves FAST diagnostic consistency and accuracy,³ however, there remain barriers in education, training, and skill maintenance that have perpetuated a lack of global pediatric FAST expertise.^{4,8,9} This lack of expertise is a primary reason for FAST not being universally implemented in pediatric clinical practice.^{10–12} Therefore, the clinical evaluation of injured children may benefit from novel methods of disseminating widespread expertise in pediatric FAST.

Deep learning is promising for clinical image processing including automatic classification of diagnostic ultrasound but has not yet been deployed for automation in pediatric

FAST.^{13,14} Deep learning anatomic ultrasound view classifiers are often the first step in developing successful models for diagnostic ultrasound applications.^{15,16} View classification allows researchers insight into the feasibility of a deep learning approach, which is essential in heterogeneous datasets.¹⁷ Inconsistency in clinician expertise and user-to-user variability of FAST adds to broad and consequential heterogeneity in image quality, and to the complexity of developing an automated view classifier.⁷ Further, deep learning models developed with heterogeneous datasets are more likely to generalize and perform within a larger range of image qualities.¹⁴ Therefore, we sought to develop and validate the accuracy of a FAST view classifier on a heterogeneous real-world dataset of pediatric FAST studies.

Materials and Methods

Study Design and Setting

This is a retrospective feasibility study of diagnostic test accuracy. We collected a convenience sample of archived pediatric FAST studies performed for a traumatic mechanism at the point-of-care by the treating clinicians during routine emergency department care at two institutions, UCSF Benioff Children's Hospital Oakland (BCHO) and UCSF Benioff Children's Hospital San Francisco (BCHSF). BCHO is a regional pediatric trauma referral center and emergency department with approximately 45,000 visits annually, and BCHSF is a tertiary referral center and pediatric emergency department receiving approximately 18,000 visits annually. Each hospital's point-of-care ultrasound quality assurance program archives FAST studies. FAST studies were included if they were interpreted by the clinician and archived in the hospital's quality assurance program. All datasets were obtained retrospectively and de-identified, with waived consent in compliance with the Institutional Review Board at the University of California San Francisco.

Data

FAST studies from March 1, 2018 through January 1, 2020 were retrieved in raw Digital Imaging and Communications in Medicine format. The dataset was secured in a high-performance computing environment in compliance with the local Institutional Review Board. All images were deidentified according to the Health Insurance Portability and Accountability Act Safe Harbor standards. All FAST studies were acquired with either the Sonosite Edge II (Fujifilm, Inc., Bothell, WA) or the Sonosite X-Porte (Fujifilm, Inc.) using phase array, linear, and curvilinear transducers with frequencies between 13 and 1 MHz. We included FAST studies for children younger than 18 years of age who were evaluated for a traumatic mechanism. Additionally, clinical and trauma registry data were linked to the respective FAST study. FAST studies were excluded if the FAST study and patient data could not be linked.

Pre-processing

A FAST study was defined as one or more video clips from a single patient and performed at a single time point. Each video clip represents one of the possible FAST views (Figure 1). FAST views include standard abdominal and thoracic sonographic views, including cardiac, pleural, abdominal upper quadrant, and suprapubic views and planes.¹⁸ The ultrasound systems capture still frames at a rate of 30 still frames per second. These FAST still

frames were all then resized to squared portable network graphics of 299×299 pixels and normalized to appear with a standardized intensity.

Ground Truth

All FAST studies were labeled by a pediatric emergency point-of-care ultrasound fellowship-trained expert with over 10 years of experience. FAST video clips were labeled once as either: cardiac, pleural, abdominal upper quadrant, or suprapubic views. The abdominal upper quadrant views also received a secondary label for laterality (corresponding to left or right abdominal upper quadrants for secondary sensitivity analysis). To ensure intra-rater reliability, 10% of studies were re-labeled. A second FAST expert with over 10 years of FAST experience independently labeled the test dataset to ensure interrater reliability. Discrepancies were resolved by rereview and then a third expert reviewer was used as a tiebreaker.

Data Partitions

FAST studies were randomly partitioned (ratio: 70:20:10) into training, validation, and test datasets.¹³ The training dataset refers to data used to fit the deep learning classifier. The validation dataset refers to data used to tune hyperparameters. The test dataset refers to an independent hold-out dataset. To maintain sample independence, FAST video clips from a single patient could be included in only one dataset.

Model Architecture and Training

The deep learning model, ResNet-152, is a convolutional neural network pre-trained on Imagenet. We choose ResNet-152 to reduce the training error and increase our likelihood of generalizability to external validation.¹⁹ A one-cycle training was implemented with a batch size of 40 still frames and a final dropout of $P = .5$. We utilized a predefined training technique including 12 epochs with 8 epochs focused on the top layer of ResNet-152, and 4 epochs allowing changes to weights across the entire network.²⁰ Spatial transformations and image manipulations were incorporated in an additional candidate model to introduce artificial noise. The data augmentation techniques included: 1) rotation (0.75 probability of rotation $0^\circ - 180^\circ$ right or left), 2) cutout (0.75 probability loss of up to four regions 10–160 pixels in size), 3) crop and resize (0.75 probability of a 299×299 pixel region selected from image scaled $0.75\times$ to $2\times$ original), and 4) adjustment of brightness and contrast (0.75 probability of a $.05\times$ to $0.95\times$ brightness and/or $0.1\times$ to $10\times$ contrast change).¹⁷

Evaluation

The primary outcome was FAST view classification accuracy by 1) video clip and 2) still frame on the test dataset. The secondary outcome evaluated view classification with the addition of differentiation of the two abdominal upper quadrant views, left and right, by video clips and still frames. Descriptive statistics were reported as frequencies and proportions to describe the sample study population and expert classification. In the test dataset, the output labels for both single still frames and full video clips were compared to the experts' manual label. For each still frame the classifier assesses a "probability" of class membership for each of the four classes. The class with the highest probability score

is assigned as the output label for a single still frame, and the video clips are assigned an output label using a plurality classification from all the still frames generated from a single video clip.¹⁶

We calculated a confusion matrix and diagnostic accuracy measures for the predicted classes. Model training and statistical analyses were performed using Python version 3.7 (Python Software Foundation) software, and R, version 3.6 (R Foundation for Statistical Computing). Class-specific accuracies are reported as the total proportion of true positives and true negatives for each predicted output class. Overall model accuracies are calculated as the total proportion of all correct classifications across all included views. Clopper–Pearson confidence intervals were also included for each proportion estimated.

Additionally, we visually represent transformed images as predicted. We selected the 4096-dimensional vector output from the last fully connected network layer (FC7) before the output prediction image. These high-dimension vectors were reduced to 50 dimensions with principal components analyses and then two dimensions with T-distributed stochastic neighbor embedding to visualize dimensionality reduction.²¹ Similarity among the model's interpretation of the images is represented by the lower Euclidian distance between the points when plotted.^{21,22} We utilized gradient-weighted class activation mapping (Grad-CAM), as a visual representation of the contribution of regions within a prediction.²³ Study experts (AK and RC) reviewed Grad-CAM predictions to assess for relevant contributions. We report the time required per training epoch and prediction run.

Results

We deidentified 701 FAST studies with recorded video clips. Two studies were excluded for missing linked patient-level data. The 699 FAST studies were included in our analysis. There were a total of 4925 video clips representing 1,062,612 still frames. The video clips were abdominal upper quadrant (45.8%), suprapubic (24.8%), cardiac (18%), and pleural (12%). The majority of FAST studies came from children less than 11 years of age, and one-third of the total studies came from children less than 6 years of age (Table 1). Sixty percent of the cohort were males. The majority of FAST studies were obtained at the level 1 pediatric trauma center and were performed by 30 attending clinicians. There was no single attending clinician that performed more than 11% of FAST studies.

The 699 FAST studies were divided into the training (70%), validation (20%), and test (10%) datasets (Table 1). Age, sex, source, and expert labeling were evenly distributed between the different datasets. The test dataset was made up of 70 FAST studies, representing 459 video clips and 104,220 still frames, which had a slightly higher proportion of studies from children 6 to 10 years of age, females, and suprapubic video clips/still frames.

FAST View Classifier

The overall classifier accuracy was 97.8% (96.0–99.0%) for video clips, and 93.4% (93.3–93.6%) for still frames (Table 2). The *F*-scores were 0.98 and 0.94 for video clips and still frames, respectively. The per-view classifier accuracies of video clips were 98.7%, 99.8%,

98.3%, and 98.9% for the cardiac, pleural, abdominal upper quadrant, and suprapubic views, respectively. The view classifier accuracies of still frames were 96.0%, 99.8%, 95.2%, and 95.9% for the cardiac, pleural, abdominal upper quadrant, and suprapubic views, respectively. The clustering analyses revealed class groupings by still frames into FAST view with minimal overlap (Figure 2). The view classifier was performed with near-identical accuracy on the augmented test dataset (Figure 3). On expert review, Grad-CAM suggested that the classifier received prediction contributions from relevant view classification sites (Figure 4). The standard clock time needed to predict the view in 1000 still frames in this set was approximately 40 seconds.

We also evaluated the accuracy of a view classifier with the additional discernment of the abdominal right and left upper quadrant views. Forty-nine (24%) video clips of the upper abdominal views received a secondary expert label for the right or left upper quadrant. A total of 151 (76%) were given a definitive label, of which 69 (46%) were right upper quadrant and 82 (54%) were left upper quadrant views. The overall view classifier accuracies were 88.5% (85.0–91.5%) for video clips and 83.5% (83.3–83.8%) for still frames (Supplemental Table 1). The F-scores were 0.88 and 0.83 for video clips and still frames, respectively. In this five-class model, view classification accuracy was lower in the abdominal upper quadrant views than the other anatomic views.

Discordant Results

The experts had five (1.1%) discordant classifications between one another with an inter-rater reliability of Fleiss Kappa of 0.98 (0.97–0.99), suggesting a near-perfect agreement between expert labels. Furthermore, eight video clips were discordant between the expert reference standard and the classifier prediction classification (Supplemental Figure 1). Four of the video clips were concordant between expert reviewers, but discordant between the classifier's prediction classification. These four discordant results were expert:classifier interpretation as cardiac:abdominal upper quadrant, abdominal upper quadrant:cardiac, and abdominal upper quadrant:suprapubic, respectively. The remaining four discordant video clips were discordant between the two experts. Three of these expert discordant video clips were concordant between expert 2 and the classifier prediction. The remaining video clip was discordant between both experts and the classifier prediction.

Discussion

We found that a deep learning pediatric FAST view classifier can perform accurately in a heterogeneous set of real-world bedside clinician-performed studies of injured children. Our study's dataset is unique in that it is a relatively large expert-labeled archive of pediatric FAST video clips and still frames from children after a traumatic mechanism. Further, the use of heterogeneous images avoids an idealized dataset which improves the generalizability of our findings. Greater than 30% of the studies are from children less than 6 years old and performed by 30 clinicians with varying levels of FAST training and expertise. Adding to the heterogeneity, FAST studies used for this study were conducted in real-time within a fast pace and high-stakes emergency environment that is prone to error.²⁴ Although the inclusion

of a heterogeneous group of FAST studies and operators adds valuable natural variation to the dataset, it increases the complexity of the classification task.

Deep learning classifiers have been successfully trained and tested in diagnostic ultrasound applications, but the feasibility of deep learning view classification for real-world pediatric FAST was unknown. Previous studies have shown echocardiography view classification and lesion detection within the liver and breast^{13,16,17} to be highly accurate. Importantly, these previous ultrasound classifiers have utilized imaging data from operators with advanced training and accreditation in image acquisition.^{15,16,25} Accredited operators are apt to derive standard sequences and higher quality image capture.^{26,27} In contrast to this, FAST is performed by the bedside treating clinician instead of a certified ultrasonography technician or other experts in image acquisition. Therefore, we conducted a feasibility study to assess whether a deep learning strategy could identify pertinent features within a FAST with multiple layers of heterogeneity.

Our findings build on previous work, demonstrating that machine learning strategies can identify pertinent features from FAST. Specifically, Sjogren et al demonstrated feasibility in using basic machine learning techniques, including support-vector machine models, in interpreting adult FAST.^{27,28} The model achieved 100% (95% confidence interval [CI], 69–100%) sensitivity and 90% (95% CI, 56–100%) specificity for free fluid detection using only the abdominal right upper quadrant view from 10 positive and 10 negative adult FAST studies. However, this study utilized a labor-intensive annotation method, a basic machine learning approach, and was performed on only a small number of FAST studies. In addition, these results may not be adaptable to pediatric FAST, where post-traumatic abdominal pathology displays wide variability and anatomic views are smaller and may vary across age groups.²⁸ In contrast, our study employed a less labor-intensive expert labeling, along with an advanced deep learning technique, which can also perform with a high degree of accuracy. Our view classification deep learning classifier also suggests that these techniques are well-suited for additional development of a multistage model to enhance the acquisition and interpretation of pediatric FAST.^{1,23,29}

Feature identification is an essential step toward developing an accurate and consistent deep learning multistage model to enhance pediatric FAST.³⁰ The FAST view classifier's high accuracy implies the ability to parse out identifying features within a pediatric FAST image. Similarly, our data augmentation simulation and Grad-CAM review suggest that a deep learning approach may be using relevant features within the image.³¹ Notably, the classifier is using pertinent features to accurately classify the different abdominal FAST views, including the suprapubic view, which is the most sensitive abdominal view for intraabdominal free fluid detection in injured children.³²

The high accuracy of our pediatric FAST view classifier suggests that a deep learning approach can achieve class prediction accuracy as well as an expert. Previous reports show that FAST has excellent consistency for view classification between experts, but only moderate agreement for less experienced FAST users.³³ Furthermore, the FAST was accepted by the trauma community in the 1990s,³⁴ however, there remains no clear association between the number of FAST studies performed and when FAST expertise is

gained.³⁵ In addition, there are barriers to training and implementation, including FAST view interpretability. For all of these reasons, it is critical that alternative methods to enhance FAST use are identified.⁷ Our classifier performed with similar accuracy to that of study experts, suggesting deep learning may offer a possible solution for poor view interpretability and view variability between clinicians with inconsistent expertise in FAST.

One of this study's goals was to develop automated processes and methods to reduce the need for human input. Human input in dataset curation is laborious, subject to error, resource-intensive, and is a rate-limiting step in high-quality, in-depth learning model development.¹³ Our automated image preprocessing pipeline and view classifier can accurately expedite the development of larger and more robust datasets for future multistage FAST models.³⁰ Our imaging preprocessing pipeline masked extraneous data outside the ultrasound viewing area to avoid training from non-clinically relevant features. Additionally, our preprocessing pipeline included data augmentation techniques to simulate image variation outside the standard differences in image quality to avoid overfitting the model. Similar to our image preprocessing, we developed a system optimized for discordant image evaluation in which experts could quickly review images (Supplemental Figure 1). This discordant image review process allows for a rapid review and understanding of model inaccuracy.

There are limitations to our study. First, the primary outcome was to develop an accurate view classifier for pediatric FAST, but not interpret the diagnostic interpretation. The recognition of relevant features within a heterogeneous dataset of pediatric FAST still frames certifies our pre-processing and gives insight into how a deep learning strategy may distinguish relevant imaging features and computational process speeds. Second, the accuracy decreased once the view classifier was tasked to discriminate between right and left abdominal upper quadrant views. The upper abdominal views are similar in sonographic appearance, which may account for the small decline in accuracy. However, it may also parallel the difficulty in the expert's interpretation. Similar to previous studies of FAST view classification, the abdominal upper quadrant views are classified together because of the similarity of features and location of free fluid collection.³³ Third, our classifier was trained using a single expert. However, previous studies and our study show a strong agreement between experts for FAST view classification.³³ Fourth, we report a single deep learning model, ResNet-152. Using this methodology, we could take advantage of efficient transfer learning, which could be scaled to future image segmentation, voxel interpretation, and may integrate into future ensemble learning strategies.

We demonstrate that our deep learning FAST view classifier has accurate view classification in heterogeneous real-world pediatric FAST video clips and still frames. Our deep learning classifier may discern relevant features within pediatric FAST images. Accurate view classification is important for quality assurance and in understanding the feasibility of developing a consistent and accurate multi-stage deep learning model for enhancing pediatric FAST.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 TR001872.

Abbreviations

BCHO, UCSF	Benioff Children's Hospital Oakland
BCHSF, UCSF	Benioff Children's Hospital San Francisco
CI	confidence interval
FAST	focused assessment with sonography for trauma
Grad-CAM	gradient-weighted class activation mapping

References

- Melniker LA, Leibner E, McKenney MG, Lopez P, Briggs WM, Mancuso CA. Randomized controlled clinical trial of point-of-care, limited ultrasonography for trauma in the emergency department: the first sonography outcomes assessment program trial. *Ann Emerg Med* 2006; 48:227–235. [PubMed: 16934640]
- Holmes JF, Kelley KM, Wootton-Gorges SL, et al. Effect of abdominal ultrasound on clinical care, outcomes, and resource use among children with blunt torso trauma: a randomized clinical trial. *JAMA* 2017; 317:2290–2296. [PubMed: 28609532]
- Kornblith AE, Graf J, Addo N, et al. The utility of focused assessment with sonography for trauma enhanced physical examination in children with blunt torso trauma. *Acad Emerg Med* 2020; 27:866–875. [PubMed: 32159909]
- Scaife ER, Rollins MD, Barnhart DC, et al. The role of focused abdominal sonography for trauma (FAST) in pediatric trauma evaluation. *J Pediatr Surg* 2013; 48:1377–1383. [PubMed: 23845633]
- Thourani VH, Pettitt BJ, Schmidt JA, Cooper WA, Rozycki GS. Validation of surgeon-performed emergency abdominal ultrasonography in pediatric trauma patients. *J Pediatr Surg* 1998; 33:322–328. [PubMed: 9498410]
- Liang T, Roseman E, Gao M, Sinert R. The utility of the focused assessment with sonography in trauma examination in pediatric blunt abdominal trauma. *Pediatr Emerg Care* 2019; 37:11.
- Jang T, Kryder G, Sineff S, Naunheim R, Aubin C, Kaji AH. The technical errors of physicians learning to perform focused assessment with sonography in trauma. *Acad Emerg Med* 2012; 19:98–101. [PubMed: 22211463]
- Steinemann S, Fernandez M. Variation in training and use of the focused assessment with sonography in trauma (FAST). *Am J Surg* 2018; 215:255–258. [PubMed: 29174769]
- Gold DL, Marin JR, Haritos D, et al. Pediatric emergency medicine Physicians' use of point-of-care ultrasound and barriers to implementation: a regional pilot study. *AEM Educ Train* 2017; 1:325–333. [PubMed: 30051051]
- Ziesmann MT, Park J, Unger BJ, et al. Validation of the quality of ultrasound imaging and competence (QUICK) score as an objective assessment tool for the FAST examination. *J Trauma Acute Care Surg* 2015; 78:1008–1013. [PubMed: 25909423]
- Moore CL, Molina AA, Lin H. Ultrasonography in community emergency departments in the United States: access to ultrasonography performed by consultants and status of emergency physician-performed ultrasonography. *Ann Emerg Med* 2006; 47:147–153. [PubMed: 16431225]

12. Ziesmann MT, Park J, Unger B, et al. Validation of hand motion analysis as an objective assessment tool for the focused assessment with sonography for trauma examination. *J Trauma Acute Care Surg* 2015; 79:631–637. [PubMed: 26402538]
13. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdom Radiol (NY)* 2018; 43:786–799. [PubMed: 29492605]
14. Shen Y-T, Chen L, Yue W-W, Xu H-X. Artificial intelligence in ultrasound. *Eur J Radiol* 2021; 139:109717. [PubMed: 33962110]
15. Liu S, Wang Y, Yang X, et al. Deep learning in medical ultrasound analysis: a review. *Engineering* 2019; 5:261–275.
16. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* 2018; 1:6. [PubMed: 30828647]
17. Arnaout R, Curran L, Zhao Y, Levine J, Chinn E, Moon-Grady A. Expert-level prenatal detection of complex congenital heart disease from screening ultrasound using deep learning. *medRxiv* 2020.
18. Richards JR, McGahan JP. Focused assessment with sonography in trauma (FAST) in 2017: what radiologists can learn. *Radiology* 2017; 283:30–48. [PubMed: 28318439]
19. Brock A, Lim T, Ritchie JM, Weston N. FreezeOut: accelerate training by progressively freezing layers. arXiv:1706.04983. *Cs Stat.* <http://arxiv.org/abs/1706.04983>. Accessed July 21, 2021.
20. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; 1:2579–2605.
21. Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. *J Digit Imaging* 2017; 30:234–243. [PubMed: 27896451]
22. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* 2020; 128:336–359.
23. Clarke JR, Spejewski B, Gertner AS, et al. An objective analysis of process errors in trauma resuscitations. *Acad Emerg Med* 2000; 7: 1303–1310. [PubMed: 11073483]
24. Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med* 2020; 3:10. [PubMed: 31993508]
25. Marriner M. Sonographer quality management. *J Echocardiogr* 2020; 18:44–46. [PubMed: 31190291]
26. Bremer ML. Relationship of sonographer credentialing to intersocietal accreditation commission echocardiography case study image quality. *J Am Soc Echocardiogr* 2016; 29:43–48. [PubMed: 26459754]
27. Sjogren AR, Leo MM, Feldman J, Gwin JT. Image segmentation and machine learning for detection of abdominal free fluid in focused assessment with sonography for trauma examinations: a pilot study. *J Ultrasound Med* 2016; 35:2501–2509. [PubMed: 27738293]
28. Simanovsky N, Hiller N, Lubashevsky N, Rozovsky K. Ultrasonographic evaluation of the free intraperitoneal fluid in asymptomatic children. *Pediatr Radiol* 2011; 41:732–735. [PubMed: 21184057]
29. Stewart J, Sprivilis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. *Emerg Med Australas* 2018; 30:870–874. [PubMed: 30014578]
30. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv:1702.08608. <http://arxiv.org/abs/1702.08608>. Accessed July 21, 2021.
31. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. arXiv:1610.02391. <https://ui.adsabs.harvard.edu/abs/2016arXiv161002391S>. Accessed October 1, 2016.
32. Brenkert TE, Adams C, Vieira RL, Rempell RG. Peritoneal fluid localization on FAST examination in the pediatric trauma patient. *Am J Emerg Med* 2017; 35:1497–1499. [PubMed: 28433455]
33. Ma O, Gaddis G, Robinson L. Kappa values for focused abdominal sonography for trauma examination interrater reliability based on anatomic view and focused abdominal sonography for trauma experience level. *Ann Emerg Med* 2004; 44:S32–S33.

34. Scalea TM, Rodriguez A, Chiu WC, et al. Focused assessment with sonography for trauma (FAST): results from an international consensus conference. *J Trauma* 1999; 46:466–472. [PubMed: 10088853]
35. Bahl A, Yunker A. Assessment of the numbers-based model for evaluation of resident competency in emergency ultrasound core applications. *J Emerg Med Trauma Acute Care* 2015; 2015:5.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

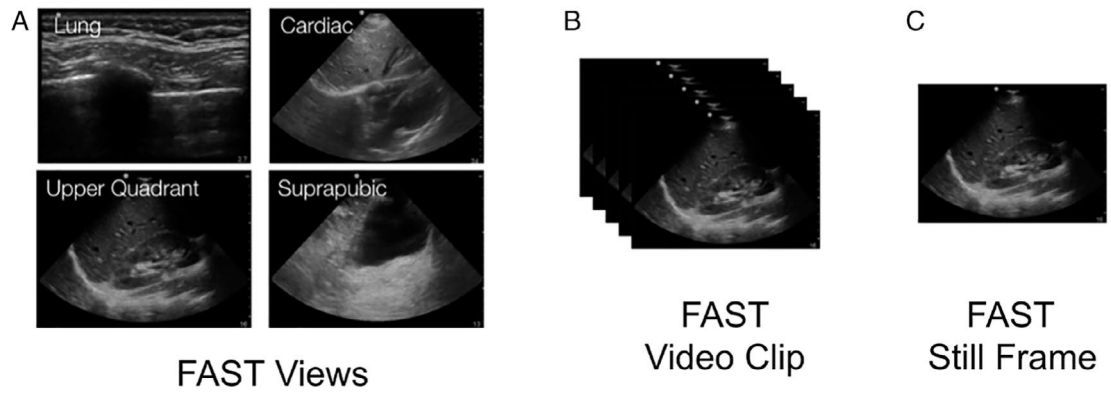


Figure 1.

A focused assessment with sonography for trauma (FAST) study is made up of a sequence of (A) FAST views. Each FAST view is archived as a (B) FAST video clip(s). Each FAST video clip is made up of multiple (c) FAST still frame.

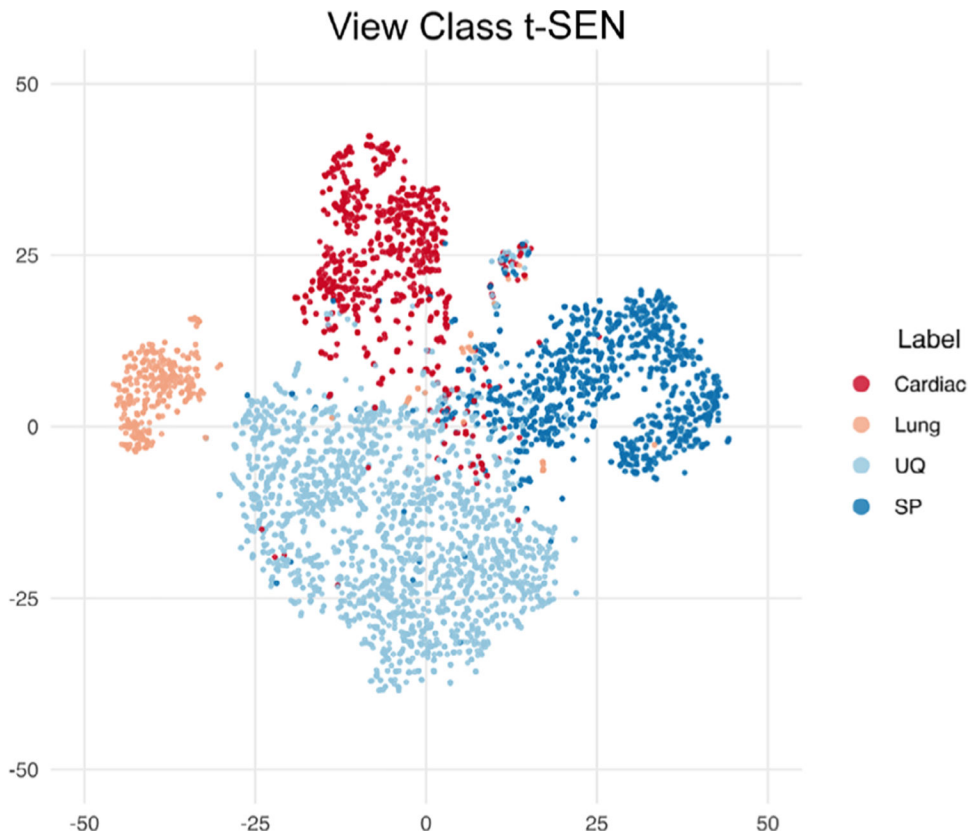


Figure 2. T-distributed stochastic neighbor embedding (t-SNE) plot. Clustering analysis of pediatric focused assessment with sonography for trauma views: cardiac views (red), pleural (peach), abdominal upper quadrant (light blue), and suprapubic views (dark blue).

Expert Label	Rotation	Cutout	Crop-Resize	Brightness-Contrast
Cardiac	93.3 (92.8–93.7)	93.4 (93.0–93.9)	94.5 (94.1–94.9)	94.6 (94.2–95.0)
Pleural	95.6 (95.1–96.1)	95.2 (94.6–95.7)	93.7 (93.1–94.3)	95.9 (95.4–96.4)
Upper Quadrant	95.5 (95.3–95.8)	96.0 (95.8–96.2)	95.5 (95.2–95.7)	95.3 (95.0–95.5)
Suprapubic	95.6 (95.3–95.9)	95.8 (95.5–96.1)	95.7 (95.4–96.0)	95.5 (95.2–95.8)
Overall	95.2 (95.0–95.3)	95.4 (95.3–95.6)	95.2 (95.0–95.4)	95.3 (95.1–95.4)

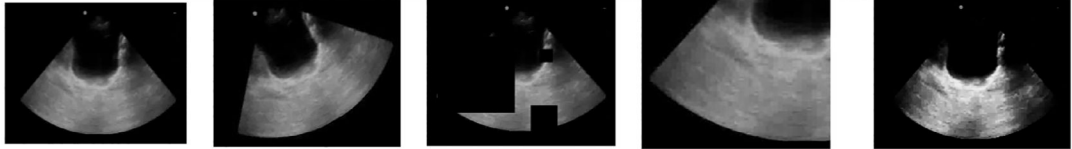


Figure 3. Pediatric focused assessment with sonography for trauma view classification accuracy (95% confidence interval) by the deep learning model for still frames after augmentation techniques.

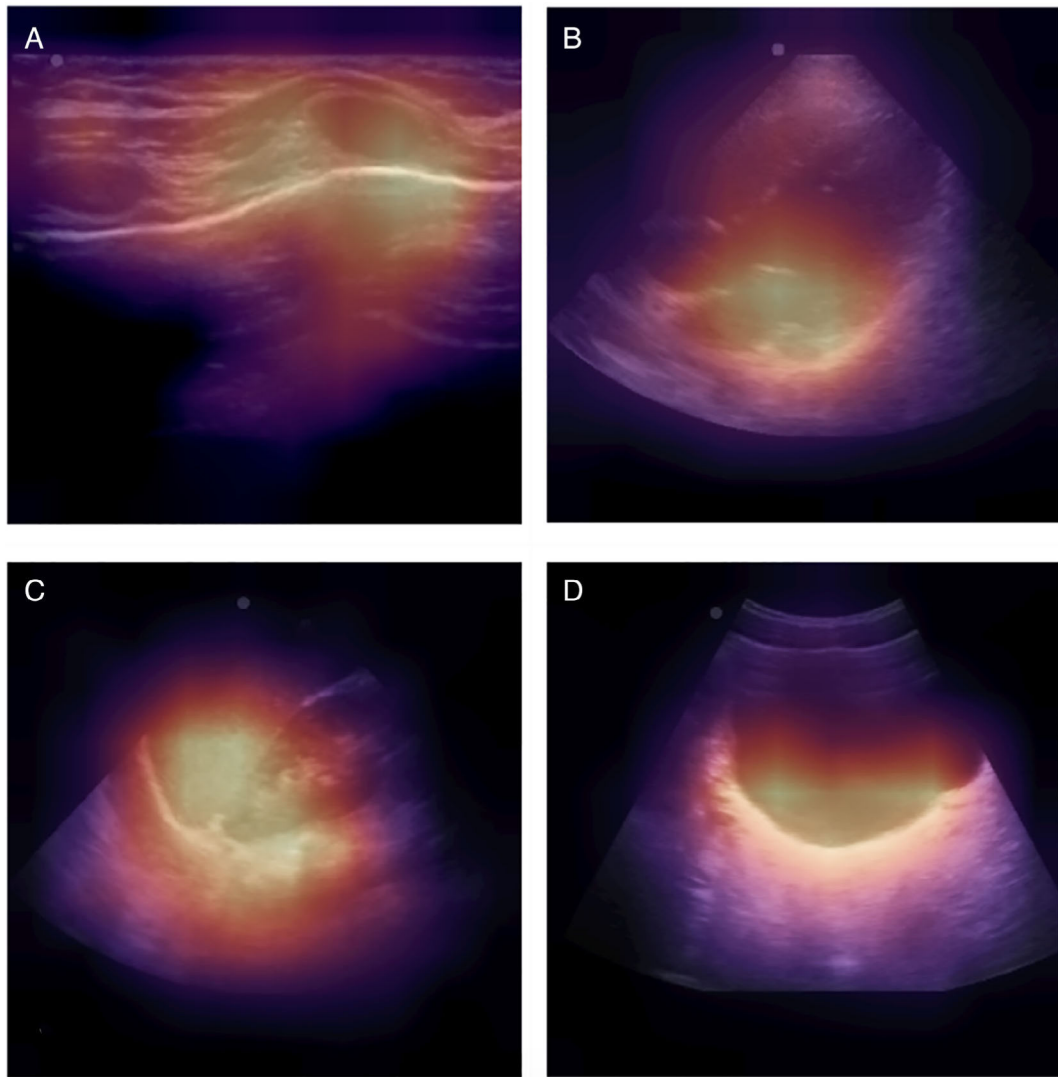


Figure 4. Gradient class activation mapping. **A**, Pleural thoracic view with the highest contribution from the costal-pleural interface. **B**, Cardiac view with the highest contribution from the left atrium. **C**, Abdominal upper quadrant view with the highest contribution from the spleen-renal-diaphragm interface. **D**, Suprapubic view with the highest contribution from the posterior bladder wall.

Table 1. Comparison of Study Sample Characteristics for Development and Testing of the Pediatric Focused Assessment With Sonography for Trauma Deep Learning Model

	Total	Training	Validation	Test
Studies	699	489	140	70
Video clips	4925	3510	956	459
Still frames	1,062,612	757,583	201,809	104,220
Age, years (studies, %)				
0–1	99	74 (15.1)	19 (13.6)	6 (8.6)
2–5	133	93 (19.0)	30 (21.4)	10 (14.3)
6–10	228	159 (32.5)	41 (29.3)	28 (40.0)
11–15	198	132 (27.1)	43 (30.7)	23 (32.9)
16–18	41	31 (6.3)	7 (5.0)	3 (4.3)
Sex (studies, %)				
Female	282	201 (41.1)	57 (40.7)	24 (34.3)
Male	417	288 (58.9)	83 (59.3)	46 (65.7)
Source (studies, %)				
Site 1	670	466 (95.3)	136 (97.1)	68 (97.1)
Site 2	29	23 (4.7)	4 (2.9)	2 (2.9)
Video clips (%)				
Still frames (%)				
Cardiac	883 (17.9)	646 (18.4)	157 (16.4)	80 (17.4)
	215,358 (20.3)	157,848 (20.8)	36,274 (18.0)	21,236 (20.4)
Pleural	565 (11.5)	407 (11.6)	104 (10.9)	54 (11.8)
	90,080 (8.5)	63,974 (8.4)	18,609 (9.2)	7497 (72)
Upper quadrant	2258 (45.8)	1614 (46.0)	444 (46.4)	200 (43.6)
	482,412 (45.4)	344,844 (45.5)	90,206 (44.7)	47,362 (45.4)
Suprapubic	1219 (24.8)	843 (24.0)	251 (26.3)	125 (27.2)
	275,762 (26.0)	190,917 (25.2)	56,720 (28.1)	28,125 (27.0)

Note: Boldface indicates video clips (%) and non-boldface indicates still frames (%).

Table 2. Pediatric Focused Assessment With Sonography for Trauma View Classification Accuracy by the Deep Learning Model for (A) Video Clips and (B) Still Frames

(A) Class-Specific Accuracy for Video Clips						
Expert Labels	Deep Learning Model Predicted Class					
	Cardiac	Pleural	Upper quadrant	Suprapubic		
Cardiac	80	0	0	0		Model overall accuracy
Pleural	0	53	0	1		97.8 (96.0–99.0)
Upper quadrant	5	0	194	1		F-Score
Suprapubic	1	0	2	122		0.978
						Brier's score
						0.378
Sensitivity	100 (94.3–100)	98.1 (88.8–99.9)	97.0 (93.3–98.8)	97.6 (92.6–99.4)		
Specificity	98.4 (96.4–99.4)	100 (98.8–100)	99.2 (96.9–99.9)	99.4 (97.6–99.9)		
Accuracy	98.7 (97.0–99.5)	99.8 (98.4–100)	98.3 (96.5–99.2)	98.9 (97.3–99.6)		
(B) Class-Specific Accuracy for Still Frames						
Expert Labels	Deep Learning Model Predicted Class					
	Cardiac	Pleural	Upper quadrant	Suprapubic		
Cardiac	19,970	12	462	792		Model overall accuracy
Pleural	28	7377	8	84		93.4 (93.3–93.6)
Upper quadrant	1989	79	43,524	1770		F-score
Suprapubic	882	13	732	26,498		0.935
						Brier's score
						0.394
Sensitivity	94.0 (93.7–94.4)	98.4 (98.1–98.7)	91.9 (91.7–92.1)	94.2 (93.9–94.5)		
Specificity	96.5 (96.4–93.6)	99.9 (99.9–99.9)	97.9 (97.8–98.0)	96.5 (96.4–96.7)		
Accuracy	96.0 (95.9–96.1)	99.8 (99.8–99.8)	95.2 (95.0–95.3)	95.9 (95.8–96.0)		

Note: Boldface indicates diagnostic test characteristics and non-boldface indicates “n” per cell.