# A quality assessment tool for focused abdominal sonography for trauma examinations using artificial intelligence

**John Cull, MD, Dustin Morrow, MD, Caleb Manasco, MD, Ashley Vaughan, MD, John Eicken, MD,**
*and* **Hudson Smith, PhD***, Greeneville, South Carolina*

**BACKGROUND:** Current tools to review focused abdominal sonography for trauma (FAST) images for quality have poorly defined grading criteria or are developed to grade the skills of the sonographer and not the examination. The purpose of this study is to establish a grading system with substantial agreement among coders, thereby enabling the development of an automated assessment tool for FAST examinations using artificial intelligence (AI).

**METHODS:** Five coders labeled a set of FAST clips. Each coder was responsible for a different subset of clips (10% of the clips were labeled in triplicate to evaluate intercoder reliability). The clips were labeled with a quality score from 1 (lowest quality) to 5 (highest quality). Clips of 3 or greater were considered passing. An AI training model was developed to score the quality of the FAST examination. The clips were split into a training set, a validation set, and a test set. The predicted scores were rounded to the nearest quality level to distinguish passing from failing clips.

**RESULTS:** A total of 1,514 qualified clips (1,399 passing and 115 failing clips) were evaluated in the final data set. This final data set had a 94% agreement between pairs of coders on the pass/fail prediction, and the set had a Krippendorff $\alpha$ of 66%. The decision threshold can be tuned to achieve the desired tradeoff between precision and sensitivity. Without using the AI model, a reviewer would, on average, examine roughly 25 clips for every 1 failing clip identified. In contrast, using our model with a decision threshold of 0.015, a reviewer would examine roughly five clips for every one failing clip — a fivefold reduction in clips reviewed while still correctly identifying 85% of passing clips.

**CONCLUSION:** Integration of AI holds significant promise in improving the accurate evaluation of FAST images while simultaneously alleviating the workload burden on expert physicians. (*J Trauma Acute Care Surg.* 2024;00: 00–00. Copyright © 2024 Wolters Kluwer Health, Inc. All rights reserved.)

**LEVEL OF EVIDENCE:** Diagnostic Test/Criteria; Level II.

**KEY WORDS:** FAST examination; ultrasound; trauma; AI.

The focused abdominal sonography for trauma (FAST) examination is a well-accepted diagnostic tool in the management of blunt trauma.[1–4] The diagnostic quality of a FAST examination is based on clear identification of certain anatomical features in each view.[5] To ensure the quality of FAST images for patient safety, continuing education, and accreditation purposes, institutions are required to review a sampling of images and videos of FAST examinations.[6]

However, this essential yet somewhat routine task of evaluating these examinations often requires the review by highly skilled sonographers and physicians. These individuals, whose expertise could be more optimally used elsewhere, engage in meticulous reviews that could benefit from a more streamlined approach. Implementing artificial intelligence (AI) for the automatic assessment of FAST examination quality would significantly economize hospital resources, presenting potential savings on a substantial scale.

To enable the training of AI for automated assessment of FAST images, a grading tool is necessary to accurately assess these images. Presently, existing criteria for grading FAST videos exhibit limitations when retrospectively evaluating ultrasound studies for quality. While some grading systems lack precise definitions, leading to poor interrater reliability, others rely on elements unrelated to the distinct identification of crucial anatomical features, such as image depth and structural orientation. This dependence on criteria not fundamentally linked to the clear identification of pivotal anatomical features might result in scoring discrepancies due to deviations from expected procedural norms.

The purpose of this study is to develop an automated assessment tool for FAST examinations using AI to implement this AI model in the FAST examination review process.

## PATIENTS AND METHODS

### Criteria Development, Trial Coding, and Criteria Refinement

A set of criteria was developed by an ultrasound fellowship–trained emergency department physician and by an experienced critical care–trained trauma surgeon to determine the image quality of FAST (IQ FAST) examinations performed at a level 1 trauma center. The IQ FAST criteria were developed from previous tools

Submitted: December 7, 2023, Revised: April 11, 2024, Accepted: May 26, 2024, Published online: September 27, 2024.

From the Department of Surgery (J.C., A.V., T.C.), and Emergency Department (D.M., C.M., J.E.), Prisma Health Upstate, Greenville, South Carolina; and Holcombe Department of Electrical and Computer Engineering (H.S.), Clemson University, Clemson, South Carolina.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.jtrauma.com).

Address for correspondence: John Cull, MD, Department of Surgery, Prisma Health Upstate, 701 Grove Rd, Greenville, SC 29605; email: john.cull@prismahealth.org.

DOI: 10.1097/TA.0000000000004425

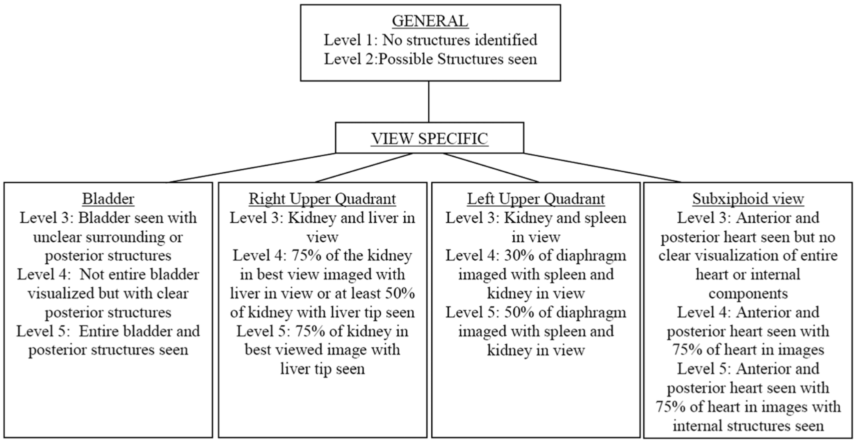
GENERAL
Level 1: No structures identified
Level 2:Possible Structures seen
VIEW SPECIFIC
Bladder
Level 3: Bladder seen with unclear surrounding or posterior structures
Level 4: Not entire bladder visualized but with clear posterior structures
Level 5: Entire bladder and posterior structures seen
Right Upper Quadrant
Level 3: Kidney and liver in view
Level 4: 75% of the kidney in best view imaged with liver in view or at least 50% of kidney with liver tip seen
Level 5: 75% of kidney in best viewed image with liver tip seen
Left Upper Quadrant
Level 3: Kidney and spleen in view
Level 4: 30% of diaphragm imaged with spleen and kidney in view
Level 5: 50% of diaphragm imaged with spleen and kidney in view
Subxiphoid view
Level 3: Anterior and posterior heart seen but no clear visualization of entire heart or internal components
Level 4: Anterior and posterior heart seen with 75% of heart in images
Level 5: Anterior and posterior heart seen with 75% of heart in images with internal structures seen

**Figure 1.** Image quality of FAST scoring criteria.

to grade FAST images deemed to be important specifically related to clinical decision making (Fig. 1).[5,7] Three trial coding rounds were performed to refine the coding criteria. Each round consisted of a labeling task wherein the creators of the scoring system used a web-based interface to review a common set of FAST clips and annotate them for quality according to the new FAST quality criteria (Fig. 2).

Four hundred thirty consecutive patient encounters from October 2017 to October 2018 were identified from all FAST images entered into the hospital's QPath database (Telexy Healthcare, Maple Ridge, BC, Canada). The video clips from these patient encounters were deidentified and uploaded to Labelbox (Labelbox, Inc., San Francisco, CA). Labelbox, a commercial data annotation tool, was used to create the labeling interface and to record annotations from coders. This tool was used for all labeling tasks presented in this work.

During round 1, coders labeled a common set of 100 clips. They agreed on whether a clip passed (quality >2) or failed (quality ≤2) 70% of the time. The coded set had a Cohen's $\kappa$ agreement score of 27% and a Krippendorff $\alpha$ score of 20%, both indicating poor agreement. After the initial labeling, clips were discussed one at a time, and a consensus label was reached for each clip. The coding criteria were updated to clarify points of disagreement that were identified while discussing the labels. During the second round, the coders labeled a new common set of 50 clips. They agreed on pass/fail 83% of the time. The set had a Cohen's $\kappa$ score of 31% and a Krippendorff $\alpha$ of 31%, indicating improvement but still unsatisfactory agreement. Consensus was reached on all clips through a follow-up review. The coding criteria were further clarified based on this discussion. During round 3, coders labeled a new common set of 50 clips. They agreed on pass/fail 89% of the time. The set had a Cohen's $\kappa$ of 69% and a Krippendorff $\alpha$ of 69%, indicating satisfactory agreement between the coders. Consensus was reached on all samples through a follow-up review. The consensus labels resulting from these three trial rounds were collected into a single "gold standard" data set containing roughly 200 clips that would serve as a reference data set for training additional coders.

## Training New Coders

After completing the three trial rounds described previously, three additional coders were trained to apply the IQ FAST criteria. Two of the coders were ultrasound fellowship–trained emergency medicine physicians, while one coder was a surgical critical care fellow. The training included one trial round in which the new coders labeled a sample (n = 50) of the criterion standard clips developed in the trial coding phase. The performance of each coder was evaluated against the criterion standard labels. All coding errors were discussed in a follow-up review for training purposes. During the training round, the three new coders annotated pass/fail quality with 80.3% accuracy, indicating strong initial performance with the new criteria. The coders then labeled an additional sample (n = 50) of the criterion standard clips. During this second training round, the three new coders annotated pass/fail quality improved to 92% accuracy.

## Full Data Set Annotation

After completing the training phase, the five coders then annotated the full set of FAST clips (n = 2,937). Each coder was responsible for a different subset of approximately 670 clips. Roughly 10% of the clips (n = 276) were labeled in triplicate to evaluate intercoder reliability. When forming the final labeled set, majority voting was used to select the final code. Clips were disqualified in the case of a tie. Nonstandard FAST examination views (such as eFAST (Extended Focused Assessment with Sonography in Trauma) views of the chest or cardiac videos in the parasternal long or parasternal short axis) were removed.

## AI Model Development

Using the large set of IQ FAST labeled clips, a deep learning-based computer vision model was trained to identify whether a clip should receive a passing or failing IQ FAST score. The Ultrasound Video Network model, which is specifically designed for ultrasound tasks, was used to evaluate the FAST labeled clips. This deep learning model uses a CNN (convolutional neural network) encoder to embed each frame of a video and an attention mechanism to aggregate relevant information from different times in the video. This approach leads to greater sample efficiency in cases where the task depends primarily on recognizing key visual features at any point in a clip as opposed to recognizing patterns in the change of visual features through time. The model design and training procedure that were used were previously presented in the original Ultrasound Video Network paper.[8] For evaluation, a random 80% to 20% train-test partition was

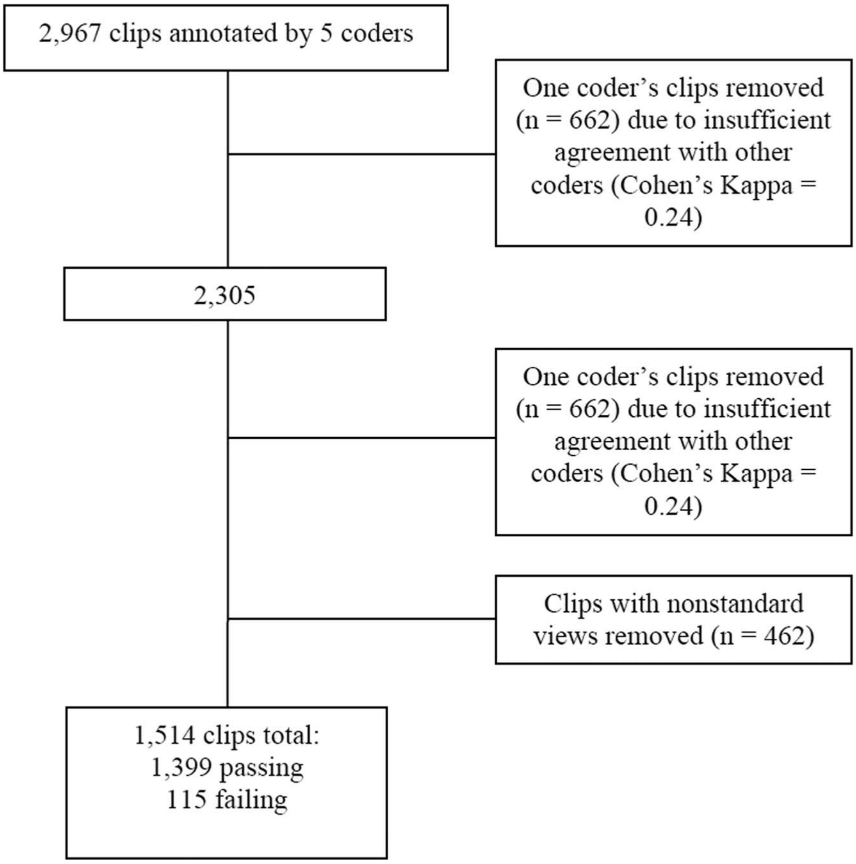
2,967 clips annotated by 5 coders
One coder's clips removed (n = 662) due to insufficient agreement with other coders (Cohen's Kappa = 0.24)
2,305
One coder's clips removed (n = 662) due to insufficient agreement with other coders (Cohen's Kappa = 0.24)
Clips with nonstandard views removed (n = 462)
1,514 clips total:
1,399 passing
115 failing

**Figure 2.** Flow chart showing number of charts reviewed and reasons omitted.

created of the data stratified on the clip quality level. Clips from a given patient appear in only the training set or test set. An ensemble of 12 models were trained on the 80% training partition following a cross-validation procedure. For the final evaluation, the average of confidence scores output was computed by each of the 12 models on the 20% test set. The averaged confidence scores were compared with the human-annotated pass/fail labels to estimate the precision, sensitivity, and specificity of the model. This article is compliant with the Clinical AI Research reporting guidelines[9] (Supplemental Digital Content, Supplementary Data 1, http://links.lww.com/TA/D921).

## RESULTS

During the study period, the average pass/fail accuracy of the coders was 91.7%. Coder 1 had disagreement in labeling quality of 17 of the 50 images and labeling 7 of the 50 views. Coder 1 had a 63% accuracy rate of identifying diagnostic quality and a 93% accuracy rate of identifying pass/fail of the images with a mean absolute error of 0.43 (Supplementary Fig. 1, http://links.lww.com/TA/D922).

Coder 2 had disagreement in labeling quality of 18 of the 50 images and labeling 12 of the 50 views. Coder 2 had a 64% accuracy rate of identifying diagnostic quality and a 94% accuracy rate of identifying pass/fail of the images with a mean absolute error of 0.50 (Supplementary Fig. 2, http://links.lww.com/TA/D923).

Coder 3 had disagreement in labeling quality of 17 of 49 images and labeling 7 of 49 views. Coder 3 had a 65% accuracy rate of identifying diagnostic quality and an 88% accuracy rate of identifying pass/fail of the images with a mean absolute error of 0.43 (Supplementary Fig. 3, http://links.lww.com/TA/D924).

All 5 coders annotated a total of 2,967 clips. It was determined that one coder did not have sufficient agreement with the other coders to label the complete set. The coder had an average Cohen's $\kappa$ of 0.24 between all other coders. This coder's samples were removed from the set before computing consensus and excluded from the labeled data set, leaving 2,305 clips. Majority voting was used to select the final code. There were 329 clips that were disqualified due to tie. Lastly, 462 clips from nonstandard views were removed, leaving 1,514 qualified clips in the final labeled data set. The set contained 1,399 passing and 115 failing clips. Among these valid clips, 245 unique clips had at least 2 annotations from different clinicians. Based on this subset, we estimated an average 94.4% agreement between pairs of coders on the pass/fail prediction. The set had a Krippendorff $\alpha$ of 66%, indicating satisfactory agreement. Precision, sensitivity, and specificity were computed for each pair of coders. This was used to compare coder agreement with model performance.This was performed by treating one coder as the ground truth and the other as the model. The roles of the coders were then swapped; the average of the two values was taken. Finally, the average of the scores across pairs of coders was obtained, weighing each pair by the number of mutually labeled samples. This resulted in an average precision of 52%, a sensitivity of 72%, and a specificity of 97%. These results represent the upper bound of the performance we could expect to obtain from an AI model.

Figure 3 quantifies the performance of the model for identifying failing clips. The model does not directly output a pass/fail prediction for a clip. Rather, it outputs a confidence score ranging from 0 to 1 with 1 being most confident that a clip is of failing quality. To make predictions, a decision threshold must be applied between 0 and 1. Tuning this threshold leads to different performance tradeoffs as shown by the curves in Figure 3. For high-decision thresholds, the model is very precise but lacks sensitivity (left side of plot). For low-decision thresholds, the model lacks precision but has high sensitivity (right side of plot). The decision threshold can be tuned to achieve the desired tradeoff between precision and sensitivity.

Table 1 examines the tradeoffs represented by the three red points from Figure 3. These decision thresholds are points along the curve shown in Figure 3 where there are no other points that have higher values of both precision and sensitivity. As the precision decreases, the sensitivity increases. Specificity levels remain relatively high even for the highest sensitivity exemplar. These results represent a dramatic improvement over naively searching for failing examinations by examining random clips with uniform probability. In that case, a reviewer would, on average, examine roughly 25 clips for every 1 failing clip identified. In contrast, using our model with a decision threshold of 0.015, a reviewer would examine roughly five clips for every one failing clip — a fivefold reduction in clips reviewed while still correctly identifying 85% of passing clips. The model at decision threshold 0.25 has somewhat better precision and specificity than the human annotators but at a lower sensitivity. From this, one can conclude that the model performs near the limit set by the consensus among the physicians.

**TABLE 1.** Analysis of Red Exemplar Points From Figure 3

| Decision Threshold | Precision | Sensitivity | Specificity |
|---|---|---|---|
| 0.25 | 58% | 43% | 99% |
| 0.015 | 19% | 81% | 85% |
| 0.0074 | 16% | 94% | 79% |
| Physician agreement | 52% | 72% | 97% |

Each row corresponds to choosing a different decision threshold and represents a different tradeoff between the precision and sensitivity with which our model identifies failing examinations. The final row marked "physician agreement" records the average precision, sensitivity, and specificity scores estimated from our physician-labeled consensus set for comparison.

## DISCUSSION

The routine collection of FAST examinations serves as a cornerstone for swift triage in trauma centers. Post hoc evaluations of these examinations are undertaken to ensure quality and to credential providers. While the 2022 Standards for the

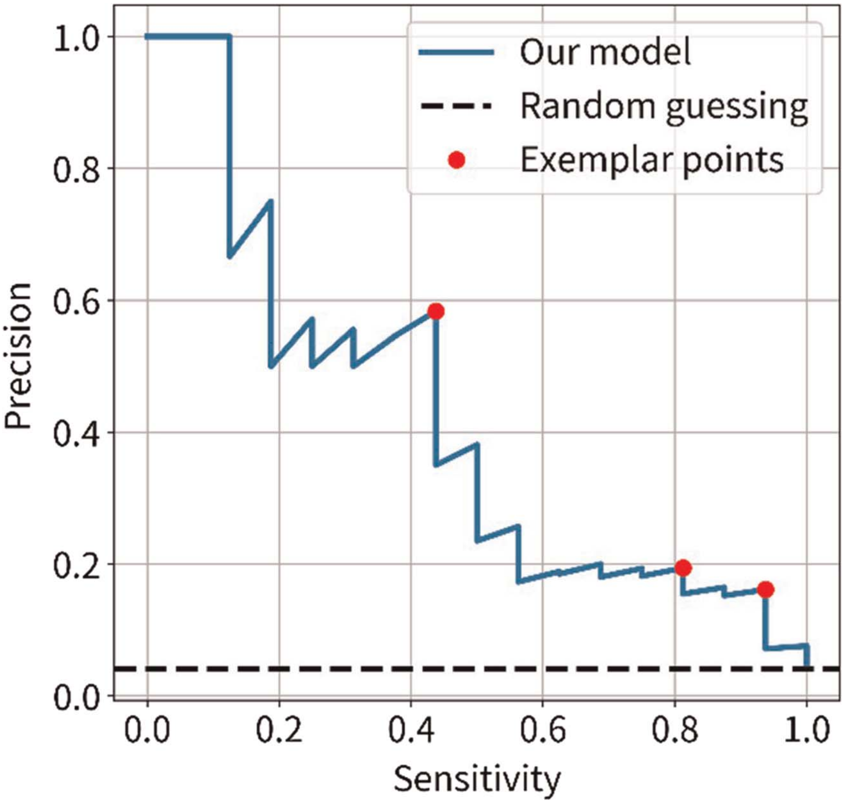

**Figure 3.** The precision of the model at increasing levels of sensitivity for the identification of failing quality clips. Precision quantifies the rate at which clips predicted failing are actually failing according to human annotators. Sensitivity quantifies the proportion of actually failing clips identified as such by the model. The curve was generated by tuning the decision threshold on model confidence scores. The horizontal dashed line represents the expected precision when randomly guessing clips are failing with probability equal to the share of failing clips in our data set. The red exemplar points represent reasonable tradeoffs between precision and sensitivity as explored further in Table 1.

Optimal Resources for the Care of Injured Patients presently do not mandate trauma programs to review ultrasound images for quality assurance, Emergency Medicine has been more vocal in advocating for bedside imaging training and quality assessment requirements. The American College of Emergency Physicians, through the Clinical Ultrasound Accreditation Program, oversees local ultrasound programs' compliance with national standards for routine review and quality enhancement.[10,11]

Currently, the predominant grading scale for assessing imaging quality is the 5-point Emergency Ultrasound Standard Reporting Guidelines (Fig. 3). However, its main drawback lies in its poor interrater reliability, mainly because of the absence of a clearly defined "minimum criteria for diagnosis" within the reporting guidelines.

Conversely, the Quality of Ultrasound Imaging and Competence (QUICk) score demonstrates good interrater reliability in evaluating the quality of ultrasound images.[7] Comprising a global rating scale and a task-specific checklist, the QUICk score assesses both the performance quality, such as probe positioning and gel application, and the task-specific elements crucial for image quality, clearly delineating the requisite ultrasound views for each FAST examination section. Nevertheless, the QUICk score was not explicitly designed for routine review and quality improvement of ultrasound images.[7,12] Some components, focusing on image orientation rather than the direct identification of anatomical features, might cause deviations from strict assessment protocols to impact assessment scores.

In our approach, we amalgamated elements from both the widely adopted Emergency Ultrasound Standard Reporting Guidelines and the anatomic features of the QUICk score to precisely define the segments of the FAST examination that meet the "minimum criteria for diagnosis." Coders, with minimal training, achieved an 80.3% accuracy in annotating pass/fail quality for each video. With subsequent rounds of training, this accuracy surpassed 90%, demonstrating strong agreement among experienced sonographers. However, accuracy declined among least experienced sonographer, potentially because of skill degradation over time or the challenges posed by evaluating numerous videos. This hybrid approach leverages the strengths of existing grading systems to establish a more comprehensive and precise framework for assessing FAST examination quality, exhibiting promising accuracy levels even with minimal training.

The academic literature and commercial domain have established a robust foundation for applying machine learning to medical imaging tasks.[13–16] These machine learning applications aim to automate traditional human expert–led medical image analyses, typically performed by professionals like sonographers or radiologists. Through a training process involving the collection of medical imagery and expert manual annotations, machine learning systems seek to approximate human expertise. These applications hold promise in reducing manual human effort required for delivering expert medical care across diverse care facilities, potentially mitigating human error.[17]

The domain of ultrasound imagery has notably attracted machine learning applications, leveraging the widespread use of ultrasound devices alongside enhanced computational capabilities to advance health care delivery. Sjogren et al.[18] successfully applied a traditional machine learning approach to segment free fluid in image frames extracted from FAST examinations. More recently, Kornblith et al.[19] used a deep learning approach for FAST view classification, providing model predictions for both static frames and video clips via aggregation. Differing from these studies, our research emphasizes the quality assessment of FAST examinations based on one or multiple ultrasound videos per patient. To our knowledge, despite precedents in related contexts,[20–23] no study has explored the application of machine learning specifically to assess the quality of individual FAST examination videos for adult trauma patients.

In our study, the performance of the trained AI model in classifying pass and fail FAST ultrasound images demonstrated comparable accuracy with that of physicians trained to grade these images using the IQ FAST criteria. Because of the limited representation of failing examinations in the data set, our AI model demonstrated a heightened level of uncertainty when categorizing examinations that fell within the margins of passing or failing grades.

In its present state, the effective utilization of this AI model necessitates a strategic approach in handling instances where the model exhibits high uncertainty in grading images. To ensure comprehensive coverage of failing data, expert reviewers would need to focus on the segments where the AI demonstrates uncertainty. Specifically, for capturing 80% of the failing data, reviewing an additional five videos is required to identify a failing video. Elevating this coverage to 94% necessitates a slightly higher review rate of approximately six additional videos for the identification of a failing video. This strategy aims to compensate for the model's uncertainty by leveraging human expertise in pinpointing the nuanced cases where the AI exhibits ambiguity, ensuring a more comprehensive and accurate assessment of failing data.

Implementation of AI to the quality review process for FAST images yields a remarkable reduction, up to 80%, in the number of videos physicians would need to review. Furthermore, continual evaluation and review of images identified as uncertain by the AI are expected to contribute to the model's ongoing refinement, potentially leading to a further reduction in the number of videos necessitating physician review. We are currently working with our institution to create the software necessary to make this AI model widely available. This new software must be developed to seamlessly interface between existing centralized ultrasound repositories while ensuring protection of patient information. Future studies should include training experienced sonographers at other institutions the IQ FAST criteria and comparing our AI model with the graded FAST images from multiple institutions.

## CONCLUSION

In conclusion, the integration of AI holds significant promise in improving the accurate evaluation of FAST images while simultaneously alleviating the workload burden on expert physicians. The next step is to develop the necessary software to incorporate the AI evaluation of images to the retrospective review process for FAST images.

### AUTHORSHIP

J.C. contributed in the literature search, study design, data collection, data interpretation, writing, and critical revision. A.V. contributed in the study design, data collection, and critical revision. C.M. contributed in the study design, data collection, and critical revision. J.E. contributed in the study design, data collection, and critical revision. D.M. contributed in the study design, data collection, and critical revision. H.S. contributed in the study design, data analysis, and data interpretation.

## ACKNOWLEDGMENTS

We used an AI-powered language model to aid in refining the language used within this article. This study was funded in part by a Research Seed Grant through Prisma Health/Clemson University. We would like to express our sincere gratitude to Dr. Thomas Coleman for his assistance in the planning and design of this study.

## DISCLOSURE

Conflicts of Interest: Author Disclosure forms have been supplied and are provided as Supplemental Digital Content (http://links.lww.com/TA/D925).

## REFERENCES

1. Ollerton JE, Sugrue M, Balogh Z, D'Amours SK, Giles A, Wyllie P. Prospective study to evaluate the influence of FAST on trauma patient management. *J Trauma*. 2006;60:785–791.
2. Teixeira PGR, Inaba K, Hadjizacharia P, Brown C, Salim A, Rhee P, et al. Preventable or potentially preventable mortality at a mature trauma center. *J Trauma*. 2007;63:1338–1337; discussion 1346-7.
3. Planquart F, Marcaggi E, Blondonnet R, Clovet O, Bobbia X, Boussat B, et al. Appropriateness of initial course of action in the management of blunt trauma based on a diagnostic workup including an extended ultrasonography scan. *JAMA Netw Open*. 2022;5:e2245432.
4. 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.
5. Liu R, Blaivas M, Moore C, Sivitz A, Flannigan M, Tirado A, et al. Emergency ultrasound standard reporting guidelines. 2018.
6. 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.
7. Ziesmann MT, Park J, Unger BJ, Kirkpatrick AW, Vergis A, Logsetty S, 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.
8. Smith DH, Lineberger JP, Baker GH. On the relevance of temporal features for medical ultrasound video recognition. Lecture Notes in Computer Science, pages 744-753, 2023;arXiv:2310.10453
9. Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundstrom C, Hedlund J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a clinical AI research (CAIR) checklist proposal. *Acta Orthop*. 2021;92:513–525.
10. Anonymous 7th Edition of the Resources for Optimal Care of the Injured Patient (2022 Standards).
11. Anonymous ultrasound guidelines: emergency, point-of-care and clinical ultrasound guidelines in medicine. *Ann Emerg Med*. 2017;69:e27–e54.
12. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84:273–278.
13. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. 2017;37:505–515.
14. Ratner M. FDA backs clinician-free AI imaging diagnostic tools. *Nat Biotechnol*. 2018;36:673–674.
15. Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, et al. Deep learning interpretation of echocardiograms. *NPJ Digit Med*. 2020;3:10.
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. doi:10.1038/s41746-1 Epub 2018 Mar 21.
17. 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.
18. 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.
19. Kornblith AE, Addo N, Dong R, Rogers R, Grupp-Phelan J, Butte A, et al. Development and validation of a deep learning model for automated view classification of pediatric focused assessment with sonography for trauma (FAST). *J Ultrasound Med*. 2022;41:1915–1924.
20. Yu F, Sun J, Li A, Cheng J, Wan C, Liu J. Image quality classification for DR screening using deep learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2017;2017:664–667.
21. Zago GT, Andreão RV, Dorizzi B, Teatini Salles EO. Retinal image quality assessment using deep learning. *Comput Biol Med*. 2018;103:64–70.
22. Lin Z, Li S, Ni D, Liao Y, Wen H, Du J, et al. Multi-task learning for quality assessment of fetal head ultrasound images. *Med Image Anal*. 2019;58:101548.
23. Taye M, Morrow D, Cull J, Smith DH, Hagan M. Deep learning for FAST quality assessment. *J Ultrasound Med*. 2023;42:71–79.