

RESEARCH

Open Access



Diagnostic accuracy of AI-assisted point-of-care ultrasound for abdominal free fluid detection in FAST trauma assessment: a systematic review and meta-analysis

Ali Çelik^{1*} , Ensar Topaloğlu² and Mümin Murat Yazıcı²

Abstract

Study objective Point-of-care ultrasound (PoCUS) is widely used in trauma care through the Focused Assessment with Sonography for Trauma (FAST) protocol, but its accuracy is highly operator-dependent. Artificial intelligence (AI) may reduce variability and improve reliability. This systematic review and meta-analysis evaluated the diagnostic accuracy of AI-assisted PoCUS for detecting free fluid in trauma patients.

Methods We searched PubMed, Scopus, and Web of Science up to April 2025, following PRISMA-DTA guidelines (PROSPERO ID: CRD420250615096). Eligible studies assessed AI-assisted PoCUS using the FAST or FAST-equivalent views and provided sufficient data for diagnostic accuracy analysis. Because one eligible study evaluated an ascites cohort rather than trauma patients, we included it in the broader abdominal free-fluid analysis but excluded it from the trauma-only sensitivity analysis. Pooled estimates were calculated with a bivariate random-effects model. Study quality was assessed using QUADAS-AI tool.

Results Seven retrospective studies ($n = 2,332$ patients, $>34,000$ images/videos) were included. The pooled analyses were based on the diagnostic units reported by the original studies, which varied across image-, frame-, video/clip-, FAST-view-, and examination-level data. In the trauma-only analysis, pooled sensitivity for detecting abdominal free fluid was 91.1% (95% CI: 77.9–96.8%) and specificity was 97.5% (95% CI: 95.3–98.7%), with negligible heterogeneity ($I^2 < 1.0$) and an AUC of 0.98. In the broader abdominal free-fluid analysis, which included one non-trauma ascites cohort, pooled sensitivity was 91.4% (95% CI: 81.6–96.3%) and specificity was 96.8% (95% CI: 86.5–99.3%), with an AUC of 0.97. CNN-based models showed similar performance (sensitivity 92.2%, specificity 95.4%, AUC 0.97). Narrative review highlighted substantial variability across models: high-performing frameworks such as YOLOv3 and ResNet50-V2 demonstrated sensitivities of 0.90–0.99, whereas others (e.g., VGG11_bn, MaskRCNN) were markedly less accurate. Evidence for pericardial effusion detection was limited to a single retrospective study and should be interpreted cautiously. No dedicated pelvic-view diagnostic accuracy data were available.

*Correspondence:

Ali Çelik
ali.celik@erdogan.edu.tr

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusion AI-assisted PoCUS shows promising retrospective diagnostic performance for detecting abdominal free fluid during FAST assessment in trauma settings. The strongest evidence currently supports abdominal applications, whereas data for the cardiac/pericardial component of the FAST examination remain insufficient. Prospective multicenter studies with real-time workflow integration are needed before routine clinical implementation can be recommended.

Trial registration PROSPERO (CRD420250615096).

Keywords Abdominal trauma, Artificial intelligence, Deep learning, Diagnostic accuracy, Hemoperitoneum, Medical image analysis, FAST, Point-of-care ultrasound, Trauma

Introduction

Trauma is a leading global cause of death and disability, especially among young adults, where rapid and accurate diagnosis can significantly impact survival. Computed tomography (CT), particularly in thoracoabdominal trauma, is widely considered the most accurate imaging modality and is routinely recommended by major medical societies such as the American College of Radiology [1]. However, ultrasound provides real-time imaging, bedside availability, and affordability, making it a valuable first-line tool, particularly for hemodynamically unstable patients [2].

The Focused Assessment with Sonography in Trauma (FAST) has become integral in early trauma evaluation, enabling rapid, non-invasive detection of intra-abdominal free fluid (AFF) and pericardial effusion (PCE), and its role has now extended beyond the emergency department into prehospital settings. According to the American Institute of Ultrasound in Medicine (AIUM) Practice Parameter, FAST provides a standardized framework for evaluating the peritoneal and pericardial spaces in trauma sonography [3]. In conventional FAST, this assessment typically includes the right upper quadrant/hepatorenal recess, left upper quadrant/splenorenal recess, suprapubic pelvic view, and cardiac/pericardial view; extended FAST (EFAST) additionally incorporates thoracic views for pneumothorax and hemothorax.

Point-of-care ultrasound (PoCUS), performed by front-line providers, allows immediate bedside assessment and has seen growing adoption across trauma care. However, its accuracy in emergency department and prehospital setting remains highly operator-dependent, limiting standardization [4–6]. Cumulative evidence from previous studies, systematic reviews, and meta-analyses indicates that FAST demonstrate high specificity but only moderate sensitivity in detecting thoracoabdominal injuries [5–10]. Some studies have suggested that integrating FAST findings with scoring systems may improve diagnostic performance in selected trauma populations [11]. Recent reviews have also emphasized the growing use of prehospital PoCUS and its feasibility for real-time or remote supervision, while underlining the need for improved standardization and interpretive consistency

[12]. Moreover, previous systematic reviews have demonstrated variable diagnostic performance of PoCUS across acute conditions, emphasizing operator dependence and underscoring the need for strategies that improve consistency in interpretation [10, 13–15].

Artificial intelligence (AI) may help address these challenges by improving consistency in ultrasound interpretation across different experience levels and by supporting automated interpretation, image acquisition, and training [16, 17]. Although research in this field has expanded rapidly, comprehensive evidence on the diagnostic performance of AI-assisted PoCUS in FAST examinations remains limited [18–21].

This systematic review and meta-analysis aimed to evaluate the diagnostic accuracy of AI-assisted PoCUS for free fluid detection during FAST assessment in trauma settings.

Materials and methods

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) framework [22]. The study protocol was prospectively registered in the PROSPERO international database (ID: CRD420250615096) to ensure transparency and methodological rigor.

Eligibility criteria

Prospective and retrospective studies that evaluated the diagnostic performance of AI-assisted PoCUS within the FAST protocol or FAST-equivalent views were included. Eligible studies were required to involve human subjects undergoing FAST or FAST-equivalent ultrasound examinations and to integrate AI or machine learning (ML) algorithms for image interpretation, guidance, or automated detection of abdominal free fluid or pericardial effusion. Studies were considered if they reported diagnostic accuracy outcomes such as sensitivity (sens), specificity (spe), likelihood ratios (LRs), diagnostic odds ratios (DOR), or if sufficient data were available to reconstruct 2 × 2 contingency tables (true positives [TP], false positives [FP], true negatives [TN], false negatives [FN]).

Although the primary objective focused on trauma-related FAST examinations, one eligible non-trauma ascites cohort using FAST-equivalent abdominal views was retained in the broader abdominal free-fluid analysis to provide a comprehensive assessment of AI-based free-fluid detection. A predefined sensitivity analysis excluding this non-trauma cohort (e.g., ascites cohorts) was conducted to determine trauma-specific diagnostic performance.

Studies were excluded if they (i) did not incorporate AI/ML in conjunction with PoCUS, (ii) were limited to animal or phantom models, (iii) focused on conditions outside the scope of trauma-related FAST, or FAST-equivalent free-fluid detection, or (iv) lacked sufficient methodological detail or outcome data. To minimize bias, case-control designs, narrative reviews, editorials, and small studies with fewer than 20 patients were also excluded. Only full-text articles published in English up to 15 April 2025 were considered; conference abstracts and other forms of grey literature were not included.

Search strategy and study selection

The detailed search strategy is provided in Table S1 (Supplement) and covered studies published up to April 2025. The following databases were systematically searched: PubMed, Scopus, Web of Science, and Google Scholar (backward and forward citation tracking). Two reviewers (MMY, ET) independently screened titles and abstracts according to the predefined eligibility criteria. Full-text articles of potentially eligible studies were then assessed in detail. Any disagreements were resolved through discussion, and a third reviewer (AÇ) was consulted when necessary.

Data extraction and quality assessment

Data from the included studies were independently extracted by two reviewers (MMY, ET). Extracted variables included study characteristics (author, year, country, and design), inclusion criteria, sample size, AI model and algorithm type, dataset composition (training/validation/test sets), validation strategy (e.g., split-sample, cross-validation, external validation), ultrasound modality and FAST views, and reference standard, and the diagnostic unit used for accuracy analysis. TP, FP, TN, and FN were extracted or reconstructed when possible, using the diagnostic unit reported by each original study, including image-, frame-, video/clip-, FAST-view-, or examination-level classifications. When patient-level 2×2 data were not available, the reported diagnostic unit from the original study was used for reconstruction and synthesis.

When multiple AI models were reported within a single study, the primary model as defined by the original study authors was selected for quantitative synthesis. If the authors explicitly identified a main or final model,

that model was included. In the absence of such designation, preference was given to models evaluated using external validation datasets; if external validation was unavailable, results derived from an independent test set were selected. Training-set performance metrics were not used for meta-analysis. Model selection was not based on highest performance metrics. If several model architectures were presented, all were reviewed narratively; however, only one prespecified primary model per study was included in the quantitative meta-analysis to maintain statistical independence of observations and avoid unit-of-analysis bias (Table S2, Supplement). Study quality was independently assessed by two reviewers using the QUADAS AI tools [23]. Disagreements were resolved through consensus, with a third reviewer (AÇ) acting as adjudicator when required.

Statistical analysis

All analyses were conducted in RStudio (version 2025.05.0 + 496, Quarto 1.6.42) using the *mada*, *metafor*, and *meta* packages. The primary synthesis applied a bivariate random-effects model (Reitsma) to estimate pooled sensitivity and specificity with 95% confidence intervals (CIs). When the bivariate model did not converge due to limited data, univariate random-effects models were used. Pooled 2×2 data were analyzed using the diagnostic units reported in the original studies; therefore, denominators in pooled analyses may represent image-, frame-, video/clip-, FAST-view-, or examination-level data rather than patient-level observations. Forest plots were generated to display study-level and pooled estimates for sensitivity, specificity, likelihood ratios (LR+, LR-), and diagnostic odds ratio (DOR). For studies that evaluated multiple AI models, only one algorithm per study was included in pooled analyses to maintain independence of data. A trauma-only sensitivity analysis was performed by excluding the non-trauma ascites cohort, and this analysis was considered the most clinically relevant estimate for trauma-related FAST performance. Overall diagnostic performance was further summarized with a hierarchical summary receiver operating characteristic (HSROC) curve and the area under the curve (AUC). Between-study heterogeneity was quantified within the bivariate framework using the Zhou and Dendukuri I^2 statistic derived from the Reitsma model [24, 25]. Because all analyses were conducted using the bivariate model, heterogeneity estimates reflect the joint variability of sensitivity and specificity. A chi-squared test ($p < 0.1$) was also applied to complement heterogeneity assessment. Publication bias was assessed using funnel plots and Deeks' asymmetry test. However, the small number of included studies ($n = 7$) limits the reliability of these analyses.

Results

Search strategy

A total of 2,066 records were identified through database searches and citation tracking. After removing duplicates and screening titles and abstracts, 43 full-text articles were assessed. Of these, 36 were excluded. Seven studies met the eligibility criteria and were included in review and quantitative synthesis. The study selection process is illustrated in Fig. 1.

Study characteristics

Details of the seven included studies are summarized in Table 1. Four studies were conducted in the United States, three in East Asia. All were retrospective in design and focused on adult patients undergoing FAST examinations primarily in trauma settings, although one study evaluated fluid detection in a non-trauma (ascites) cohort using FAST-equivalent abdominal views; this study was retained in the broader abdominal free-fluid analysis and excluded from trauma-only sensitivity analyses [26]. Sample sizes ranged from 20 to 864 patients, with a total of 2,332 patients and over 34,000 ultrasound images or videos across studies. Across studies, diagnostic accuracy was reported using different units of analysis, including image-, frame-, video/clip-, FAST-view-, or examination-level classifications. Six studies employed POCUS-based FAST protocols focused on the right upper quadrant (RUQ) [16, 18–20, 27, 28], while two also incorporated left upper quadrant (LUQ) [16, 26] and one included a pericardial view [29]. No included study reported a dedicated AI-based diagnostic accuracy analysis for the suprapubic pelvic FAST view.

Artificial intelligence models varied across studies and included deep learning architectures such as Inception v3, U-net, ResNet50-V2, YoloV3 (including Darknet-53), and AutoML-based models, alongside traditional machine learning methods such as support vector machines. Among these, convolutional neural network (CNN)-based approaches were the most frequently used. Most of the studies employed internal validation strategies, including hold-out sets, 5-fold or 10-fold cross-validation, and stratified sampling. One study (Jeong et al.) conducted external validation using an independent institutional dataset in addition to training and test sets [28]. Manual expert-based annotations were used as the reference standard in all studies, with classification provided by experienced emergency physicians, surgeons, radiologist or multidisciplinary panels.

Quality assessment

Figure 2 summarizes the risk of bias assessment based on the QUADAS-AI tool. The reference standard and flow and timing domains were consistently rated as low risk, with no applicability concerns across studies. However,

five studies showed high or unclear risk in patient selection, mainly due to retrospective design and unclear source population. Risk of bias in the index test domain was also frequently rated as unclear due to limited reporting on model development and validation. Publication bias was not detected by funnel plot (Figure S1, supplement) and Deeks' test ($p = 0.116$).

Diagnostic performance of PoCUS

Table 2 and the forest plots in Fig. 3 present the pooled diagnostic performance metrics of AI-assisted PoCUS for detecting abdominal free fluid, pericardial free fluid, or overall FAST positivity, defined as the presence of abdominal and/or pericardial fluid. The denominators reported in these analyses refer to the diagnostic units used in the original studies, including image, frame, video/clip, FAST view, or examination-level data, rather than uniformly patient-level observations.

For abdominal free-fluid detection, pooled sensitivity for identifying abdominal free fluid ranged from 91.1% to 93.4%, while specificity ranged from 96.8% to 97.5%. The highest sensitivity was observed in the subgroup focused specifically on the right upper quadrant (RUQ) (93.4%, 95% CI: 83.1–97.6%). Specificity in this subgroup reached 96.9% (95% CI: 94.2–98.4%). In the clinically most relevant trauma-only abdominal free-fluid analysis, after exclusion of the non-trauma ascites cohort ($n = 7,997$), pooled sensitivity was 91.1% (95% CI: 77.9–96.8%) and specificity increased to 97.5% (95% CI: 95.3–98.7%), with negligible heterogeneity ($I^2 < 1.0$) and an AUC of 0.984 (95% CI: 0.983–0.986). In the broader abdominal free-fluid analysis, which included the non-trauma ascites cohort ($n = 8,317$), pooled sensitivity was 91.4% (95% CI: 81.6–96.3%) and specificity was 96.8% (95% CI: 86.5–99.3%), with an AUC of 0.970 (95% CI: 0.949–0.981) and moderate heterogeneity ($I^2 = 57.0$).

Across all pooled analyses, pooled positive likelihood ratios (LR^+) ranged from 18.94 to 36.47, while pooled negative likelihood ratios (LR^-) remained consistently low (0.07–0.09), suggesting favorable likelihood-ratio profiles in the retrospective datasets. The corresponding diagnostic odds ratios (DOR) ranged from 245.0 to 446.1, with the highest DOR observed in the RUQ subgroup (446.1; 95% CI: 243.01–812.68).

The analysis of “any FAST view” ($n = 8,393$) demonstrated pooled sensitivity of 91.1% (95% CI: 82.7–95.7%) and specificity of 96.2% (95% CI: 87.3–98.0%), with moderate heterogeneity ($I^2 = 47.5$) and an AUC of 0.968 (95% CI: 0.946–0.979). In the corresponding trauma-only analysis, excluding the ascites cohort ($n = 8,073$), sensitivity remained 91.1% and specificity increased to 97.1%, with heterogeneity decreasing to $I^2 < 1.0$ and AUC rising to 0.983 (95% CI: 0.981–0.985) (Fig. 4).

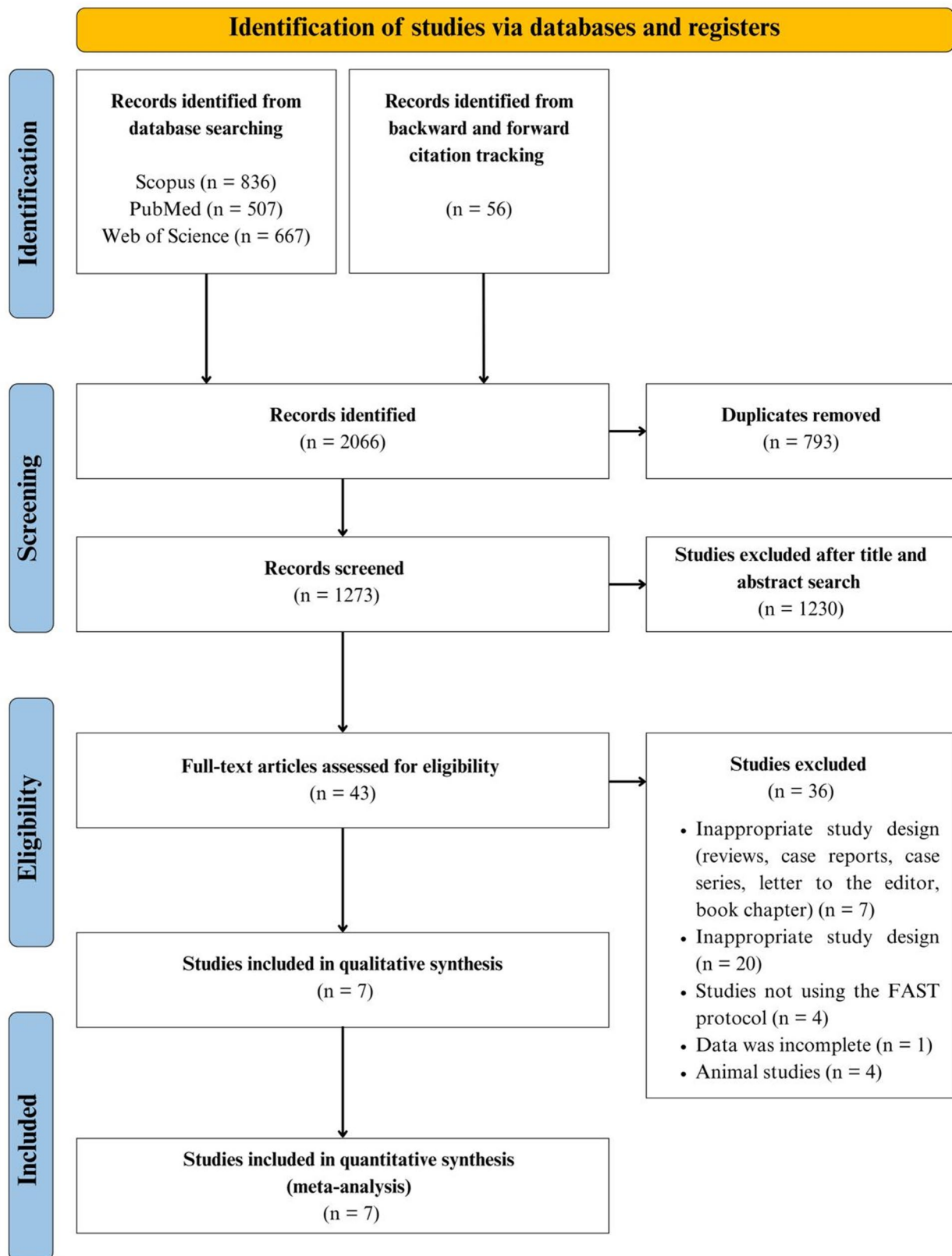
**Fig. 1** Flowchart of the study selection

Table 1 Characteristics of included studies

Author and Year	Country	Study Design	Inclusions	Population	Model	Test Datasets	Validation	Imaging and View	Reference Standard	Algorithm
Sjogren 2016 [19]	USA	Retrospective	Adult patients (≥ 18 years) from 2020–2021	20 patients, 20 US videos, 1264 images	SVM	823 images	Test and training sets divided – IV	POCUS – RUQ	Manual annotation by EP based on visual evidence of free fluid	Traditional ML
Cheng 2021 [27]	Taiwan	Retrospective	Adult patients (≥ 18 years) from 2020–2021	324 patients, 22,368 US images	ResNet50-V2	6118 images	Test, validation and training sets – IV	POCUS – RUQ	Consensus decision of 3 out of 4 EU	CNN - Classification
Lin 2022 [26]	China	Retrospective	US images from ascites patients (2016–2017)	845 patients; 2985 US images	U-net	320 images	Test and training sets divided – IV	POCUS – RUQ, LUQ	Manual annotation by EP based on visual evidence of free fluid	CNN - Segmentation
Lew 2023 [16]	USA	Retrospective	Adult trauma patients (2015–2022)	109 patients, 245 US videos, 6,608 US images	Inception v3	762 images	Test and training sets divided – IV	POCUS – RUQ, LUQ	Expert review by S or R based on visualization of anatomic landmarks	CNN - Classification
Leo 2023 [18]	USA	Retrospective	Adult patients (≥ 18 years) from 2010–2018	94 patients, 94 US videos	YoloV3	94 videos	Test and training sets divided – IV	POCUS – RUQ	Confirmation of hemoperitoneum via CT imaging or operative report	CNN - Detection
Jeong 2023 [28]	Republic of Korea	Multicenter Retrospective	Adult trauma patients (2019–2021)	864 patients, 2200 US images	Auto-ML	200 images	External validation	POCUS – RUQ	Consensus classification by three EME with over 10 years of FAST experience	Hybrid -AutoML
Yildiz Potter 2023 [29]	USA	Retrospective	Adult patients from 2010–2018	76 patients, 76 US videos	YoloV3- Darknet-53	76 videos	Test and training sets divided – IV	POCUS – PV	Manual expert segmentation and classification confirmed by EME per ACEP guidelines	CNN - Detection

Abbreviations: EP: Experienced Physician, S: Surgeon, R: Radiologist, US: Ultrasound, EU: Experienced Ultrasonographer, ACEP: American College of Emergency Physicians, POCUS: Point-of-Care Ultrasound, FAST: Focused Assessment with Sonography for Trauma, RUQ: Right Upper Quadrant, LUQ: Left Upper Quadrant, PV: Pericardial View, CT: Computed Tomography, CNN: Convolutional Neural Networks,

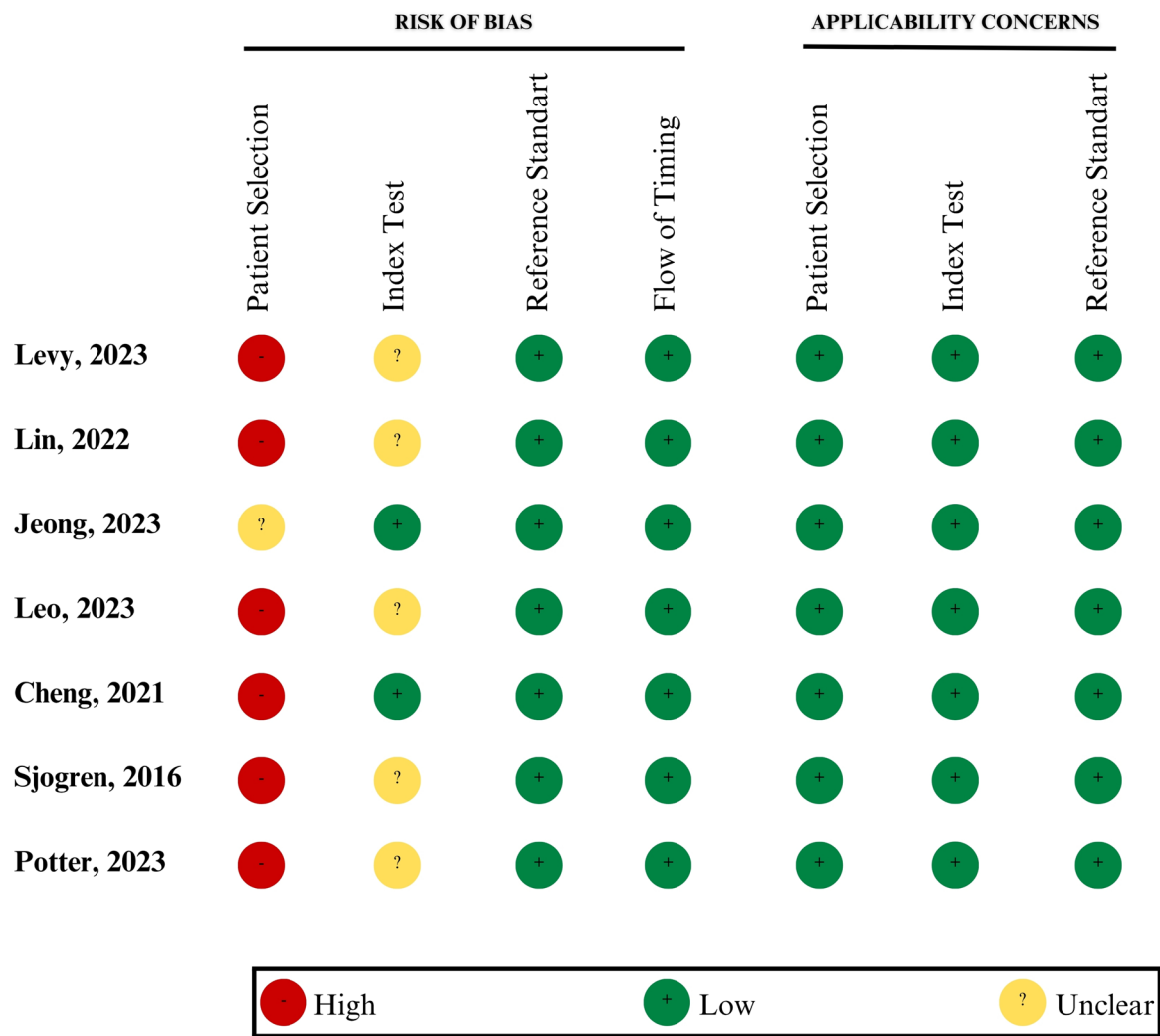
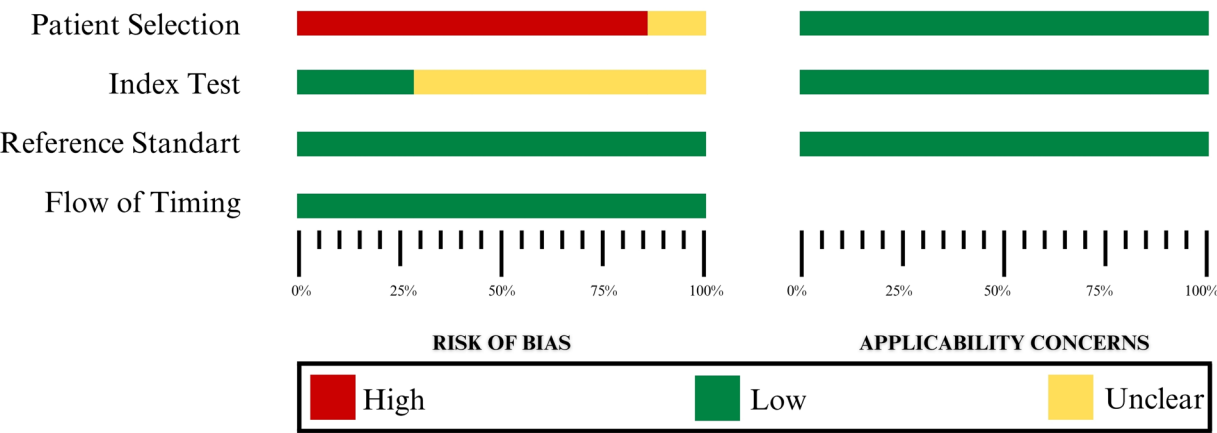


Fig. 2 Quality assessment with QUADAS-AI

Table 2 Pooled diagnostic accuracy of AI-assisted PoCUS for detecting free fluid using the FAST protocol in trauma settings

Metric (95% CI)	Abdominal free fluid (n = 8317)	Abdominal free fluid* (n = 7997)	Free fluid in RUQ (n = 7235) &	Any free fluid (any FAST view) (n = 8393)	Any free fluid (any FAST view)* (n = 8073)	Any free fluid with CNN based PoCUS* (n = 7050)	Any free fluid with CNN based PoCUS (n = 7370)
Sensitivity	0.914 (0.816–0.963)	0.911 (0.779–0.968)	0.934 (0.831–0.976)	0.911 (0.827–0.957)	0.911 (0.808–0.961)	0.922 (0.792–0.974)	0.928 (0.839–0.969)
Specificity	0.968 (0.865–0.993)	0.975 (0.953–0.987)	0.969 (0.942–0.984)	0.962 (0.873–0.980)	0.971 (0.955–0.982)	0.954 (0.943–0.962)	0.951 (0.751–0.992)
LR+	28.61 (7.00–126.53)	36.47 (20.49–59.00)	30.23 (15.54–50.28)	23.86 (7.33–83.49)	31.61 (21.26–44.30)	19.93 (16.73–23.02)	18.94 (3.82–114.98)
LR-	0.08 (0.04–0.19)	0.09 (0.03–0.22)	0.07 (0.02–0.17)	0.09 (0.05–0.18)	0.09 (0.04–0.19)	0.08 (0.03–0.22)	0.08 (0.04–0.17)
DOR	323.54 (91.40–1159.54)	401.22 (240.37–663.49)	446.10 (243.01–812.68)	258.47 (85.72–812.55)	345.07 (206.04–572.34)	244.95 (88.85–678.99)	249.01 (52.91–1258.78)
AUC	0.970 (0.949–0.981)	0.984 (0.983–0.986)	0.984 (0.982–0.986)	0.968 (0.946–0.979)	0.983 (0.981–0.985)	0.965 (0.959–0.970)	0.965 (0.934–0.979)
Heterogeneity	$I^2 = 57$	$I^2 \approx 0$	$I^2 \approx 0$	$I^2 = 47.5$	$I^2 \approx 0$	$I^2 \approx 0$	$I^2 = 53.8$

Abbreviations: LR+: positive likelihood ratio; LR-: negative likelihood ratio; DOR: diagnostic odds ratio; 95% CI: 95% confidence interval; *: trauma-only analysis (Lin et al. excluded); RUQ: right upper quadrant; CNN: convolutional neural networks; AUC: area under the curve; &: free fluid in RUQ estimates. AUC and 95% CIs were derived from the HsROC (Reitsma) model using parametric bootstrap (10,000 simulations)

In analyses restricted to CNN-based PoCUS models, diagnostic performance remained promising in retrospective datasets. For the broader CNN cohort ($n=7,370$), pooled sensitivity was 92.8% (95% CI: 83.9–96.9%) and specificity was 95.1% (95% CI: 75.1–99.2%), with moderate heterogeneity ($I^2 = 53.8$) and an AUC of 0.965 (95% CI: 0.934–0.979). In trauma-only CNN analyses ($n=7,050$), heterogeneity was negligible ($I^2 < 1.0$), with pooled sensitivity of 92.2% (95% CI: 79.2–97.4%) and specificity of 95.4% (95% CI: 94.3–96.2%), yielding an AUC of 0.965 (95% CI: 0.959–0.970).

Discussion

This meta-analysis suggests that AI-assisted point-of-care ultrasound (PoCUS) has promising retrospective performance for detecting abdominal free fluid during FAST assessment, although the evidence for the pericardial component remains limited. Across abdominal free-fluid analyses, pooled sensitivity ranged from 91.1% to 93.4%, and specificity ranged from 96.8% to 97.5%. These findings suggest a favorable diagnostic profile in the available retrospective datasets, supported by likelihood ratios (LR+: 18.94–36.47; LR-: 0.07–0.09), high diagnostic odds ratios (DOR: 244.95–446.10), and AUC values above 0.96 across analyses. No included study provided a dedicated AI-based diagnostic accuracy analysis for the pelvic FAST view.

When the analysis was restricted to trauma cohorts, heterogeneity became negligible ($I^2 < 1.0$) and the AUC increased to 0.984. This finding suggests that at least part of the variability seen in the broader analysis may be related to differences in patient populations, particularly the inclusion of the non-trauma ascites cohort. CNN-based models also performed well in the retrospective datasets, with sensitivities of 92%–93% and specificities of 95%–96%, although performance still varied across model architectures. The remaining heterogeneity in some pooled estimates may reflect differences in AI architecture, training datasets, image labeling standards, and validation methods. Overall, these findings support the feasibility of AI-enhanced PoCUS for abdominal free fluid detection in emergency and trauma settings, but they should be interpreted cautiously given the retrospective nature of the evidence and the limited data for the cardiac/pericardial view.

These results are in line with broader evidence showing that deep learning achieves high diagnostic performance across diverse imaging modalities, where AI-based classification frequently matches or exceeds expert-level interpretation [17]. In FAST, however, diagnostic performance remains closely tied to operator experience, image acquisition quality, and interpretation skills. Previous studies have shown that FAST generally has high specificity but limited sensitivity for intra-abdominal free fluid, even

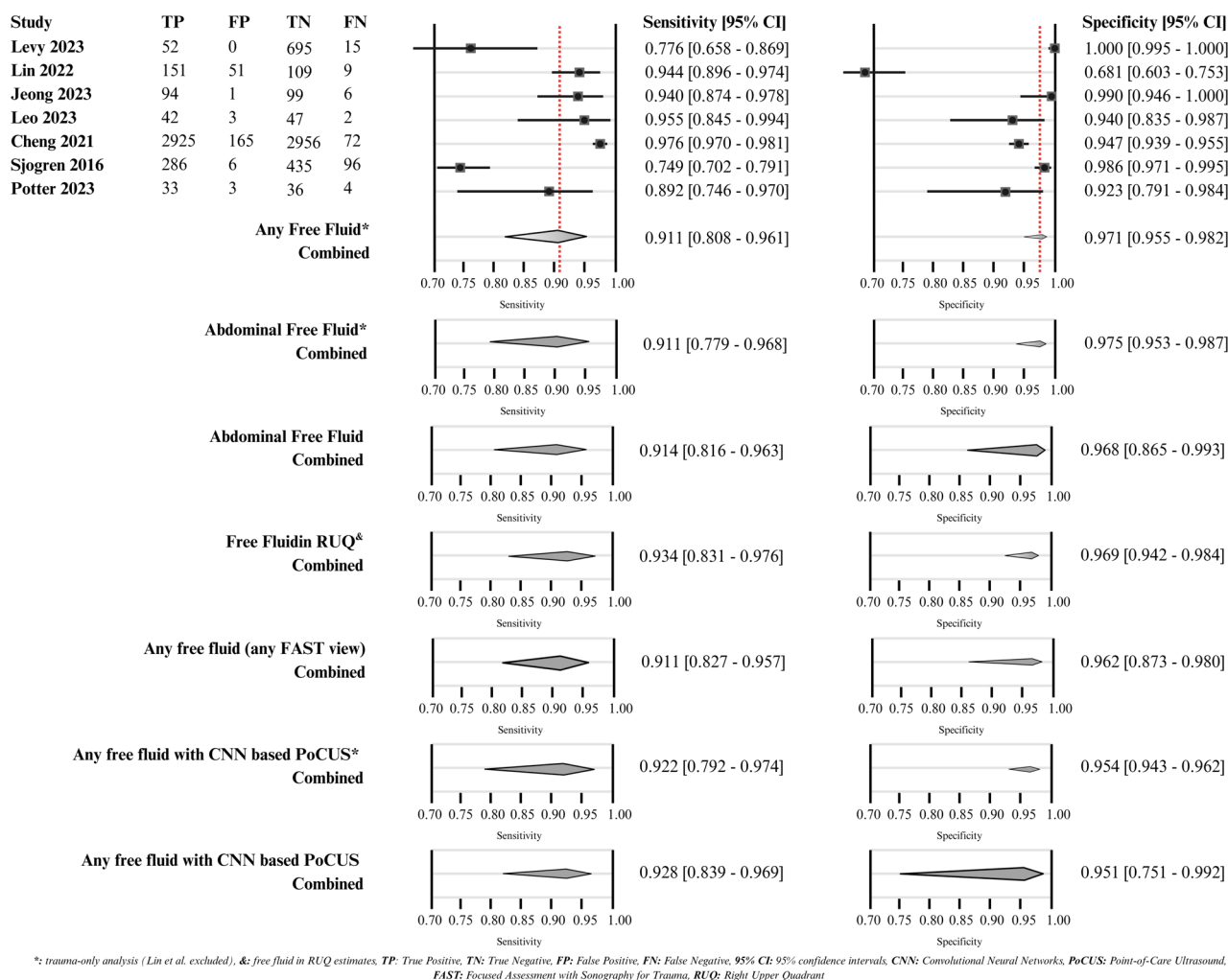
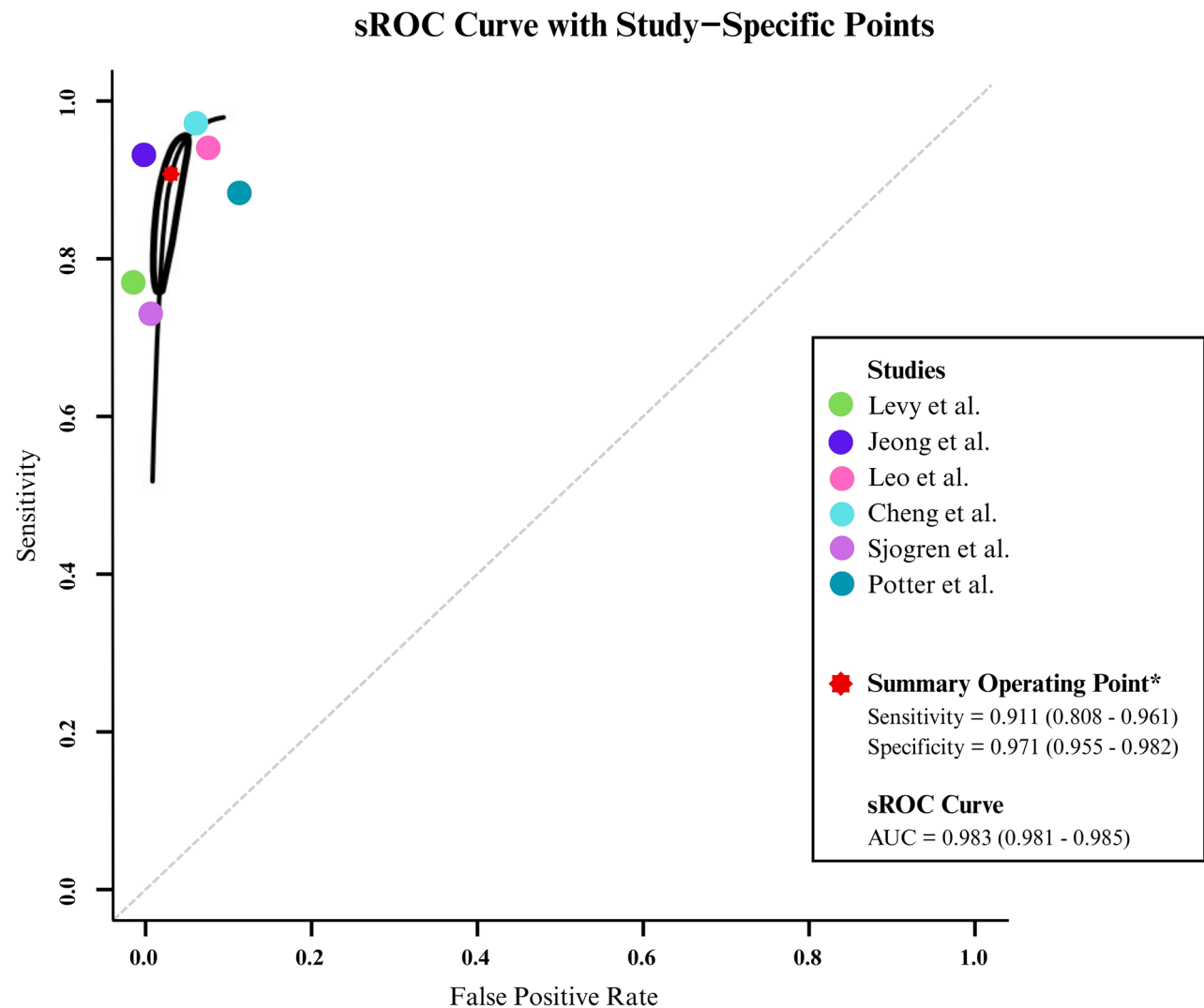


Fig. 3 Forest plots of diagnostic accuracy metrics

when performed by trained emergency physicians [5, 6, 8, 30]. Against this background, AI may be most useful as a supportive tool that improves consistency, assists less experienced users, and reduces variability in image interpretation, rather than as a standalone diagnostic method. The rapid growth of AI and deep learning research in PoCUS, also reflected in recent bibliometric analyses, further supports the clinical interest in this field [31].

Taken together, the pooled estimates suggest that AI may help standardize FAST interpretation, particularly for abdominal free fluid detection. The near-zero heterogeneity observed in the trauma-only analyses is encouraging, but it should not be interpreted as proof of clinical stability, given the small number of retrospective studies and the use of pre-acquired datasets. CNN-based models appeared to perform consistently well, although accuracy still varied across model architectures and FAST views. Overall, the current evidence supports AI as an adjunct to PoCUS interpretation rather than as a replacement for clinician judgment.

When examined narratively, model-specific performance for abdominal FAST positivity revealed substantial variation across studies. Earlier machine-learning approaches, such as the support vector machine model evaluated by Sjögren et al. (2016), supported the feasibility of automated FAST interpretation, but sensitivity remained modest despite very high specificity [19]. With the evolution of deep learning architectures, convolutional neural networks (CNNs) achieved clear improvements. For example, ResNet50-V2, evaluated by Cheng et al. on more than 5400 test images and over 6000 validation images, achieved near-perfect sensitivity (0.99) with strong specificity (0.91) [27]. U-Net demonstrated heterogeneous performance: Lin et al. [26] reported high sensitivity (0.94) but relatively low specificity (0.68), whereas Jeong et al. [28] enhanced U-Net with AutoML-based modifications, achieving a more balanced profile with external validation sensitivity of 0.94 and internal validation sensitivity of 0.95, alongside specificity approaching 0.99. Similarly, Leo et al. (2023) demonstrated that



*: trauma-only analysis (Lin et al. excluded), sROC: Summary receiver operating characteristics, AUC: Area under the curve, PoCUS: Point-of-Care Ultrasound, FAST: Focused Assessment with Sonography for Trauma

Fig. 4 Hierarchical summary receiver operating curve (HsROC)

YOLOv3 performed robustly on video-based datasets, combining high sensitivity (0.96) with high specificity (0.94) [18]. However, in the same study, alternative video-based methods such as MaskRCNN and ResNet architectures produced markedly poorer results, with sensitivity as low as 0.38–0.59. The largest dataset to date was reported by Levy et al. (2023), comprising 245 videos (6608 images, 762 test cases) [16]. In this study, InceptionV3 achieved notable performance with sensitivity of 0.78, perfect specificity (1.00), and an exceptionally high DOR (4711). DenseNet121 demonstrated excellent specificity (0.96) but lower sensitivity (0.72), while ResNet50 provided a more balanced profile (sensitivity 0.90, specificity 0.86). By contrast, the VGG11_bn architecture evaluated in the same dataset showed very poor sensitivity and only moderate specificity, highlighting the

wide variability in model performance even when trained and tested on the same dataset, a variability that is likely amplified by the predominant reliance on internal validation across most studies.

For pericardial effusion, the published evidence is limited to a single retrospective study by Yıldız Potter et al. (2023) [29]. In that work, the YOLOv3-Darknet53 model showed the most consistent performance, with a sensitivity of 0.89, a specificity of 0.92, and a diagnostic odds ratio close to 78. By contrast, other approaches such as U-Net, MaskRCNN, and SSD were far less reliable. Given that cardiac FAST data are derived from a single-center retrospective cohort, these findings should be interpreted cautiously and cannot yet support broad generalization. Larger, multicenter, and prospective studies will be essential before AI-assisted interpretation of

the pericardial FAST view can be considered for clinical implementation.

Clinical implications

Based on our findings, AI-assisted PoCUS using the FAST protocol can be considered a promising adjunct for the detection of intra-abdominal free fluid in trauma patients, particularly when rapid decision-making is critical. In positive cases, the combination of PoCUS with AI support may increase diagnostic confidence when used together with clinical assessment and conventional PoCUS interpretation.

However, a negative or indeterminate AI-assisted PoCUS should not be used in isolation to exclude intra-abdominal or pericardial injury; further imaging, clinical reassessment, or surgical consultation remains essential to ensure patient safety. For pericardial effusion, the evidence is currently too limited to support definitive conclusions, and AI should be regarded as supportive rather than standalone until validated by larger, multicenter prospective studies. Importantly, beyond the emergency department, AI-assisted PoCUS may have particular relevance in prehospital and mass-casualty settings, where its potential to standardize image interpretation and accelerate triage could be useful when operator experience varies. However, this remains an area for future research rather than an established clinical application. Given the near-zero heterogeneity observed in trauma-restricted analyses, AI integration may be potentially useful in environments where operator experience is variable. It should be noted, however, that all included studies evaluated pre-acquired images or videos in retrospective settings, and none assessed fully automated real-time AI integration within clinical workflow. Therefore, while the diagnostic performance appears promising, the direct impact of AI-assisted PoCUS on clinical decision-making, workflow efficiency, or patient outcomes remains unknown. This concept aligns with emerging tele-supervision models demonstrating that remote real-time guidance of prehospital PoCUS is technically feasible and effective in maintaining image quality and diagnostic reliability [32]. However, prospective studies are needed to determine whether similar benefits can be achieved with AI-assisted interpretation in real-world trauma care.

Limitations

Although our findings suggest promising retrospective diagnostic performance of AI-assisted FAST, several limitations must be acknowledged. First, most included studies relied on internal validation, typically dividing datasets into training and test sets, and in some cases into training, validation, and test sets. These approaches, while useful for model development, may inflate performance estimates due to limited dataset diversity and the

absence of true external validation across independent populations. Therefore, the generalizability of current results to real-world clinical settings remains uncertain. Moreover, nearly all studies were limited to 2D ultrasound imaging, without volumetric or 3D data that could provide more robust assessment of free fluid. Certain FAST regions, particularly the left upper quadrant and pelvis, were underrepresented, further constraining the scope of current evidence. In particular, no included study provided a dedicated AI-based diagnostic accuracy analysis for the pelvic FAST view, and the pericardial component was represented by only one retrospective study.

Second, heterogeneity across AI architectures, validation strategies, and dataset compositions represents an additional limitation. Although trauma-restricted analyses demonstrated minimal statistical heterogeneity, this finding should be interpreted cautiously given the limited number of included studies and shared methodological characteristics across datasets. Differences in image acquisition protocols, labeling standards, and preprocessing pipelines may still influence pooled estimates. Furthermore, near-zero heterogeneity estimates in some subgroup analyses may reflect statistical shrinkage within small meta-analytic samples rather than true clinical or methodological uniformity.

Third, in studies evaluating multiple AI architectures, only a single prespecified primary model per study was included in the quantitative synthesis to maintain statistical independence and avoid unit-of-analysis bias. Although model selection followed a predefined hierarchy (external validation preferred over independent test sets), the exclusion of alternative architectures may limit the ability to fully capture performance variability within individual datasets. This approach, while methodologically necessary, may have reduced within-study variability and potentially influenced pooled precision estimates. In addition, pooled 2×2 data were derived from mixed diagnostic units, including image-, frame-, video/clip-, FAST-view-, or examination-level classifications, rather than uniformly patient-level observations. This may have increased the apparent precision of the pooled estimates, especially when multiple observations originated from the same patient or examination.

Fourth, all included studies were retrospective, introducing potential selection bias and limiting capture of the full spectrum of trauma presentations encountered in routine practice. Most were conducted in single centers and adult populations, restricting generalizability to pediatric patients and to diverse healthcare systems, particularly in low-resource settings.

Contrary to experimental or phantom-based feasibility studies, all datasets included in the pooled analysis were derived from human subjects; however, several

investigations were conducted under controlled research conditions rather than embedded within routine clinical workflow.

Fifth, transparency in model development and reporting was limited in several studies (e.g., preprocessing steps, hyperparameter tuning), which hinders reproducibility and increases the risk of bias.

Finally, none of the included studies evaluated real-time automated integration of AI-assisted FAST within emergency department or prehospital workflows. The diagnostic performance observed in retrospective, pre-acquired image datasets does not necessarily translate into improved acquisition quality, faster decision-making, or better patient outcomes in real-world trauma care. Prospective, multicenter studies incorporating real-time deployment and workflow integration are required before routine clinical implementation can be recommended.

Conclusion

AI-assisted PoCUS using the FAST protocol shows promising retrospective diagnostic performance for detecting abdominal free fluid during FAST assessment in trauma settings. These findings suggest that AI may serve as a useful supportive tool in emergency and surgical settings, particularly where rapid image interpretation is critical. However, the available evidence remains limited. Most studies were retrospective and relied largely on internal validation, and real-time integration into clinical workflows has not yet been prospectively evaluated. Evidence for the cardiac/pericardial component of the FAST examination is especially scarce, and no dedicated pelvic-view evidence was identified. Prospective, multicenter studies with real-time workflow integration are needed before AI-assisted FAST can be recommended for routine trauma care.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12873-026-01616-6>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

None.

Author contributions

The initials of the contributing authors are listed in brackets after the relevant parts of the research: literature search (AC, MMY), study design (AC, MMY), legislative applications (AC), data collection (MMY, ET), supervision and quality control (AC), statistical data analysis (AC), data interpretation (AC), and drafting of the manuscript (AC, MMY, ET). All authors were involved in the writing and critical revision of the manuscript and approved the final version. AC take responsibility for the entire paper.

Funding

This study has been supported by the Recep Tayyip Erdoğan University Development Foundation (Grant number: 02026003011179).

Data availability

The data used in this systematic review and meta-analysis were extracted from previously published studies. All data generated or analyzed during this study are included in this published article and its supplementary materials.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Previous presentation

An abstract of this study was accepted for presentation at The Ultrasound Event | 2026 AIUM Annual Convention, Philadelphia, Pennsylvania, USA, May 27–30, 2026.

Human rights

The principles outlined in the Declaration of Helsinki were followed.

Registration

We formally registered the protocol in the PROSPERO database, ID: CRD420250615096, (Clinical trial number: not applicable).

Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT as a tool to refine the language, focusing on correcting typographical and grammatical errors. All edits and revisions were carefully reviewed and finalized by the authors, who take full responsibility for the content and accuracy of this publication.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Emergency Medicine, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye

²Department of Emergency Medicine, Recep Tayyip Erdoğan University Training and Research Hospital, Rize, Türkiye

Received: 2 March 2026 / Accepted: 11 May 2026

Published online: 16 May 2026

References

1. Shyu JY, Khurana B, Soto JA, Biffi WL, Camacho MA, Diercks DB, et al. ACR Appropriateness Criteria® Major Blunt Trauma. *J Am Coll Radiol*. 2020. <https://doi.org/10.1016/j.jacr.2020.01.024>. 17.
2. Natarajan B, Gupta PK, Cemaj S, Sorensen M, Hatzoudis GI, Forse RA. FAST scan: is it worth doing in hemodynamically stable blunt trauma patients? *Surgery*. 2010;148. <https://doi.org/10.1016/j.surg.2010.07.032>.
3. The AIUM practice parameter for the performance of the extended. Focused assessment with sonography for trauma (EFAST). *J Ultrasound Med*. 2023;42. <https://doi.org/10.1002/jum.16027>.
4. Harel-Sterling M, Kwan C, Pirie J, Tessaro M, Cho DD, Coblenz A, et al. Competency standard derivation for point-of-care ultrasound image interpretation for emergency physicians. *Ann Emerg Med*. 2023;81. <https://doi.org/10.1016/j.annemergmed.2022.11.002>.
5. Netherton S, Milenkovic V, Taylor M, Davis PJ. Diagnostic accuracy of eFAST in the trauma patient: A systematic review and meta-analysis. *Can J Emerg Med*. 2019;21. <https://doi.org/10.1017/cem.2019.381>.
6. Gamberini L, Scquizzato T, Tartaglione M, Chiarini V, Mazzoli CA, Allegri D, et al. Diagnostic accuracy for hemoperitoneum, influence on prehospital times and time-to-definitive treatment of prehospital FAST: a systematic review and individual participant data meta-analysis. *Injury*. 2023;54. <https://doi.org/10.1016/j.injury.2023.03.024>.
7. Stengel D, Leisterer J, Ferrada P, Ekkernkamp A, Mutze S, Hoenning A. Point-of-care ultrasonography for diagnosis thoracoabdominal injuries in patients

- with blunt trauma. *Cochrane Database Syst Reviews*. 2018;2018. <https://doi.org/10.1002/14651858.CD012669.pub2>.
8. Akoglu H, Celik OF, Celik A, Ergelen R, Onur O, Denizbasi A. Diagnostic accuracy of the Extended Focused Abdominal Sonography for Trauma (E-FAST) performed by emergency physicians compared to CT. *Am J Emerg Med*. 2018. <https://doi.org/10.1016/j.ajem.2017.11.019>.
 9. Buyurgan ÇS, Yarkac A, Bozkurt S, Köse A, Akverdi Akar G, Mete R, et al. Diagnostic accuracy of E-FAST examination performed by newly trained emergency physicians and its impact on clinical outcomes. *BMC Emerg Med*. 2025;25. <https://doi.org/10.1186/s12873-025-01386-7>.
 10. Lin K, Te, Lin ZY, Huang CC, Yu SY, Huang JL, Lin JH, et al. Prehospital ultrasound scanning for abdominal free fluid detection in trauma patients: a systematic review and meta-analysis. *BMC Emerg Med*. 2024;24. <https://doi.org/10.1186/s12873-023-00919-2>.
 11. Nasr Isfahani M, Nasri Nasrabadi E, Rabiei Z, Fatemi NAS, Heydari F. Enhancing clinical risk assessment in pediatric blunt abdominal trauma: a novel scoring system using ultrasound and laboratory data. *BMC Emerg Med*. 2025;25. <https://doi.org/10.1186/s12873-025-01196-x>.
 12. Hellenthal KEM, Porschen C, Whent J, Lange M. Evolving role of point-of-care ultrasound in prehospital emergency care: a narrative review. *Scand J Trauma Resusc Emerg Med*. 2025;33. <https://doi.org/10.1186/s13049-025-01443-x>.
 13. Kim DJ, Burton JE, Hammad A, Sabhaney V, Freder J, Bone JN, et al. Test characteristics of ultrasound for the diagnosis of peritonsillar abscess: a systematic review and meta-analysis. *Acad Emerg Med*. 2023;30. <https://doi.org/10.1111/acem.14660>.
 14. Çelik A, Topaloğlu E, Yazici MM. Diagnostic accuracy of point-of-care ultrasound for diverticulitis: a systematic review and meta-analysis. *BMC Emerg Med*. 2026;26. <https://doi.org/10.1186/s12873-025-01431-5>.
 15. Çelik A, Hamdioğlu E. Test characteristics of emergency physician-performed point-of-care ultrasound for the diagnosis of intrauterine pregnancy: a systematic review and meta-analysis. *Intern Emerg Med*. 2026. <https://doi.org/10.1007/s11739-025-04255-0>.
 16. Levy BE, Castle JT, Virodov A, Wilt WS, Bumgardner C, Brim T, et al. Artificial intelligence evaluation of focused assessment with sonography in trauma. *J Trauma Acute Care Surg*. 2023. <https://doi.org/10.1097/TA.0000000000004021>.
 17. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *npj Digit Med*. 2021;4. <https://doi.org/10.1038/s41746-021-00438-z>.
 18. Leo MM, Potter IY, Zahir A, Vaziri A, Jung CF, Feldman JA. Using deep learning to detect the presence and location of hemoperitoneum on the focused assessment with sonography in trauma (FAST) examination in adults. *J Digit Imaging*. 2023;36. <https://doi.org/10.1007/s10278-023-00845-6>.
 19. 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. <https://doi.org/10.7863/ultra.15.11017>.
 20. Chiu IM, Lin CHR, Yau FFF, Cheng FJ, Pan HY, Lin XH, et al. Use of a deep-learning algorithm to guide novices in performing focused assessment with sonography in trauma. *JAMA Netw Open*. 2023;6. <https://doi.org/10.1001/jamanetworkopen.2023.5102>.
 21. Çelik A, Yazici MM. The advancement of artificial intelligence in point-of-care ultrasound (POCUS): a bibliometric analysis. *Sonography*. 2026;13. <https://doi.org/10.1002/sono.70023>.
 22. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, MacAskil P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ*. 2020;370. <https://doi.org/10.1136/bmj.m2632>.
 23. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat Med*. 2021;27. <https://doi.org/10.1038/s41591-021-01517-0>.
 24. Frank RA, Salameh JP, Islam N, Yang B, Murad MH, Mustafa R, et al. How to critically appraise and interpret systematic reviews and meta-analyses of diagnostic accuracy: a user guide. *Radiology*. 2023;307. <https://doi.org/10.1148/RADIOLOGY.221437>.
 25. Zhou Y, Dendukuri N. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: The case of meta-analyses of diagnostic accuracy. *Stat Med*. 2014;33. <https://doi.org/10.1002/sim.6115>.
 26. Lin Z, Li Z, Cao P, Lin Y, Liang F, He J, et al. Deep learning for emergency ascites diagnosis using ultrasonography images. *J Appl Clin Med Phys*. 2022;23. <https://doi.org/10.1002/acm2.13695>.
 27. Cheng CY, Chiu IM, Hsu MY, Pan HY, Tsai CM, Lin CHR. Deep learning assisted detection of abdominal free fluid in morison's pouch during focused assessment with sonography in trauma. *Front Med (Lausanne)*. 2021;8. <https://doi.org/10.3389/fmed.2021.707437>.
 28. Jeong D, Jeong W, Lee JH, Park SY. Use of automated machine learning for classifying hemoperitoneum on ultrasonographic images of morison's pouch: a multicenter retrospective study. *J Clin Med*. 2023;12. <https://doi.org/10.3390/jcm12124043>.
 29. Yildiz Potter İ, Leo MM, Vaziri A, Feldman JA. Automated detection and localization of pericardial effusion from point-of-care cardiac ultrasound examination. *Med Biol Eng Comput*. 2023;61. <https://doi.org/10.1007/s11517-023-02855-6>.
 30. Yazici MM, Yavaş Ö, Çelik A, Altuntaş G, Altuntaş M, Bilir Ö, et al. The role of repeated extended FAST in patients with stable blunt thoracoabdominal trauma. *Ulusal Travma ve Acil Cerrahi Dergisi*. 2023;29. <https://doi.org/10.14744/tjtes.2022.93529>.
 31. Yazici MM, Yavaş Ö. The development of point-of-care ultrasound (POCUS): Worldwide contributions and publication trends. *J Clin Ultrasound*. 2024. <https://doi.org/10.1002/jcu.23846>.
 32. Hermann M, Hafner C, Scharner V, Hribersek M, Maleczek M, Schmid A, et al. Remote real-time supervision of prehospital point-of-care ultrasound: a feasibility study. *Scand J Trauma Resusc Emerg Med*. 2022;30. <https://doi.org/10.1186/s13049-021-00985-0>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.