

REVIEW ARTICLE

Artificial Intelligence as a Second Reader for Focused Point-of-Care Ultrasound (POCUS) in the Emergency Department: A Scoping Review

Milad Bakhshi¹, Saba Aghajani¹, Alireza Majidi^{1*}

1. Department of Emergency Medicine, Shohadaye Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Received: August 2026; Accepted: September 2026; Published online: October 2026

Abstract: **Introduction:** Artificial intelligence (AI)/ machine learning (ML) as a second reader for focused point-of-care ultrasound (POCUS) may address two key emergency department (ED) challenges: operator variability and workflow delays. This review maps the existing literature on the application of AI/ML as a second reader for POCUS in the ED. **Methods:** We included studies on AI/ML for POCUS interpretation, acquisition guidance, or quality assessment in the ED/urgent care settings, with AI as a second reader supporting clinical decisions. We excluded fully autonomous AI, non-emergency settings, healthy volunteers, phantom data, and non-English publications. Two reviewers independently screened records from five databases (PubMed/MEDLINE, Scopus, Embase, Web of Science, IEEE Xplore; 2020– August 17 2026), extracting data on study characteristics, POCUS application, AI role, performance metrics, technology readiness level (TRL), and clinical implementation status. **Results:** Of 1,884 records identified, 44 studies met inclusion criteria. Lung ultrasound was the most frequent application (38.6%). AI's primary role was as an interpretation assistant (77.3%). Models demonstrated high diagnostic performance (e.g., sensitivities $\geq 95\%$ for hemoperitoneum and DVT detection), with one study achieving 100% sensitivity and specificity for pneumothorax detection, significantly outperforming 11 expert clinicians. However, most studies were at a preclinical validation stage (TRL: 4–5; 56.8%), with only 18 studies (40.9%) testing prototypes in live clinical workflows. Major limitations identified were small, single-center datasets, reliance on expert-acquired images, and lack of investigation into critical real-world challenges like domain shift and automation bias. **Conclusions:** AI for POCUS in the ED shows strong diagnostic performance in controlled research settings but remains distant from robust clinical integration. The field is characterized by a predominance of early-stage validation studies. Future research must prioritize multi-center, multi-device external validation, inclusion of novice-generated data, prospective workflow integration, and explicit safety evaluations to translate technical promise into reliable clinical impact.

Keywords: Artificial Intelligence; Ultrasonography; Point-of-Care Systems; Emergency Service; Hospital; Triage; Scoping Review

Cite this article as: Bakhshi M, Aghajani S, Majidi A. Artificial Intelligence as a Second Reader for Focused Point-of-Care Ultrasound (POCUS) in the Emergency Department: A Scoping Review. Arch Acad Emerg Med. 2026;14(1):e60. <https://doi.org/10.22037/aaem.v14i1.3030>.

1. Introduction

Rapid, accurate diagnosis in the Emergency Department (ED) directly affects patient outcomes in a high-acuity, time-pressured environment. Increasing patient volumes and crowding continue to strain clinical resources, creating a pressing need for tools that accelerate diagnostic workflows without compromising accuracy (1, 2). In this context, focused Point-of-Care Ultrasound (POCUS) has become an indispensable extension of the physical exam. Protocols like E-FAST (Extended Focused Assessment with Sonography for Trauma), and focused cardiac, lung, and vascular ultrasounds, allow clinicians to rapidly answer binary, life-threatening questions at the bedside: Is there free fluid? Is cardiac activity present? Are there signs of pneumothorax or pulmonary edema? And etc. (3, 4).

However, the widespread adoption and consistent utility of POCUS face significant barriers. Proficiency requires substantial training and practice to achieve both technical skill in image acquisition and cognitive skill in interpretation (5). This variability poses a particular challenge in busy emergency departments (EDs) where clinicians of varying expertise must make swift decisions, potentially leading to diagnostic errors

or the underutilization of this powerful tool (6). Furthermore, even for experts, the integration of POCUS into a chaotic triage workflow can be time-consuming, potentially offsetting its speed benefits (5).

Deep learning (DL) has shown expert-level accuracy in detecting and quantifying pathologies in medical imaging (7, 8). This capability may address POCUS challenges in emergency care. Artificial intelligence (AI) is not meant to replace clinicians but to serve as an intelligent "second reader". This role can manifest in three primary ways: (1) Automated Interpretation, providing real-time detection and highlighting of pathological findings (e.g., pneumothorax, B-lines) (9); (2) Acquisition Guidance, offering real-time feedback to guide a user, especially a novice, in obtaining a diagnostic view (10); and (3) Image Quality Assessment, automatically determining if a clip is interpretable, thereby enabling efficient workflow or remote expert review (11).

Despite a surge in technical research, the integration of AI for POCUS in the specific context of ED remains an evolving and fragmented field. Existing literature spans multiple disciplines, with studies varying widely in clinical focus, stage of validation, and proposed integration into clinical workflow. While recent reviews discuss the promise of AI in ultrasound broadly (12), there is a lack of a systematic mapping of the evidence specifically focused on the triage and immediate assessment

*Corresponding Author: Alireza Majidi; Shohadaye Tajrish Hospital, Shahrday Avenue, Tajrish Square, Tehran, Iran. Email: Alirezamajidi330@yahoo.com; ORCID: 0000-0002-8732-140x; Tel: 009522721155.

phase in emergency care, where decision-making is most time-critical and the need for decision support is most acute.

This scoping review aimed to systematically map the existing literature on the application of AI/ML as a second reader for POCUS during the triage and immediate assessment phase in the ED. By clarifying the current landscape, active research frontiers, and persistent gaps, this review aims to inform future research priorities and guide the development of clinically impactful solutions for emergency care.

2. Methods

2.1. Study design and setting

We conducted and reported this scoping review following the PRISMA extension for scoping reviews (PRISMA-ScR) (13). We prepared a protocol a priori describing objectives, eligibility criteria, and methods. Studies were selected using a Population-Concept-Context (PCC) framework. Detailed PCC and main search string are provided in Table 1.

2.2. Research Questions

1. What are the key clinical applications of AI/ML as a second reader for POCUS in ED triage?
2. What are the primary roles of AI (interpretation assistance, acquisition guidance, quality assessment) in this context?
3. What is the current Technology Readiness Level (TRL) of these AI systems, and what evidence exists for their clinical implementation?
4. What are the major evidence gaps and limitations that prevent translation of AI POCUS tools from research settings to routine clinical practice?

2.3. Information Sources and Search Strategy

We searched five databases: PubMed/MEDLINE, Scopus, Embase, Web of Science, and IEEE Xplore from January 1, 2020 to August 17, 2026. Reference lists of included articles and selected reviews were hand-searched. The core search string combined terms for AI & machine learning (ML) and POCUS and ED & Triage (Table 1). Full, database-specific search strings are provided in Supplementary S1. The search period was chosen to capture recent AI advances, as deep learning applications in POCUS have proliferated since 2020.

2.4. Eligibility Criteria

Studies were selected based on the following eligibility criteria, organized using the Population-Concept-Context (PCC) framework:

Population: Studies conducted on human patients in an emergency or urgent care setting or using datasets derived from these settings.

Concept/Intervention: The AI model must analyze raw or processed POCUS images or video clips as a second reader supporting human clinical decision-making. Outcomes must include diagnostic accuracy, image quality assessment, impact on clinical decision-making/triage, or workflow efficiency.

Context: ED or similar urgent care setting where immediate clinical decisions are made.

We excluded studies where AI was fully autonomous without human-in-the-loop; AI used for non-emergency comprehensive ultrasounds; AI used for guiding procedures (e.g., vascular

access); image compression/archival without diagnostic intent; case reports, editorials, letters, conference abstracts, narrative reviews, protocols; non-English publications; studies published before 2020 unless seminal foundational works. Full details of inclusion/exclusion criteria are provided in supplementary S2.

2.5. Selection of sources of evidence

Records were imported into EndNote and screened. After removing the duplicate records, two reviewers independently screened titles/abstracts and full texts, resolving disagreements through discussion or a third reviewer. Details for each step are reported in the PRISMA-ScR (13) flow diagram (Figure 1).

2.6. Data charting process

Data were extracted independently by two reviewers using a standardized data extraction form that was pilot tested on five studies and refined through team discussion. Data charting was performed independently, with disagreements resolved by consensus or consultation with a third reviewer. The following variables were sought:

- **Study characteristics:** author, year, country, study design (prospective, retrospective, mixed/feasibility)
- **Population characteristics:** sample size (patients and/or clips), patient demographics, clinical setting
- **POCUS application:** organ/system (lung, cardiac, vascular, FAST/trauma, musculoskeletal, inferior vena cava (IVC)/carotid)
- **AI/ML methodology:** algorithm type, training approach, input data format, model architecture
- **AI role:** interpretation assistant, acquisition guide, quality assessor
- **Performance metrics:** sensitivity, specificity, area under the curve (AUC), accuracy, F1-score, kappa agreement, with 95% confidence intervals where available
- **Technology readiness level (TRL):** assessed using standardized criteria (1-9 scale) adapted from NASA guidelines (14)
- **Clinical implementation:** whether tested in live clinical workflow (Yes/No), implementation setting
- **Validation approach:** internal vs. external validation, cross-validation methods
- **Limitations:** sample size constraints, single-center bias, operator dependence, domain shift issues

2.7. Synthesis of results

As this is a scoping review, we did not perform quantitative meta-analysis or statistical pooling of results. Instead, findings were synthesized narratively using thematic synthesis, grouping studies by POCUS application domain, AI role, and TRL. When multiple studies addressed similar clinical tasks, we reported ranges of performance metrics to illustrate variability rather than calculating pooled estimates. Consistency across studies was assessed qualitatively, noting whether findings were generally consistent, whether contradictory results existed, and potential reasons for inconsistency.

2.8. Quality and risk of bias assessment

We assessed the methodological quality of included studies using criteria adapted for scoping reviews: study design (prospective/retrospective/feasibility), sample size adequacy, validation approach (internal vs. external), performance reporting completeness (presence of 95% confidence intervals), and data source representativeness. Formal risk-of-bias tools (e.g., QUADAS-2, PROBAST) were not used as these are designed for diagnostic accuracy systematic reviews and are not appropriate for scoping reviews. We critically appraised each study to contextualize findings and identify potential biases affecting the cumulative evidence, including publication bias, selective reporting, and incomplete outcome reporting. A summary of methodological limitations is provided in the Results section, and their implications are discussed in the Discussion section.

3. Results

3.1. Characteristics of included studies

The database search yielded 1,884 records. After duplicate removal, 1,353 studies underwent title and abstract screening. 136 studies proceeded to full-text selection, of which 75 were excluded. In total, 44 studies met all inclusion criteria and were incorporated into the final synthesis (Figure 1).

Nearly half of the included studies were conducted prospectively (21, 47.7%), 16 studies (36.4%) were retrospective, and the remaining 7 studies (15.9%) used mixed or alternative designs such as feasibility assessments, secondary analysis of existing clinical trial data, and retrospective analysis of prospectively gathered datasets.

A detailed summary of study characteristics, POCUS applications, AI roles, performance metrics, clinical implementation, and Technology Readiness Level for all 44 included studies is provided in Table 2.

Across the included studies, lung ultrasound was the most frequently examined POCUS application (17, 38.6%), followed by cardiac imaging (9, 20.5%) and vascular assessments such as DVT evaluation (7, 15.9%). FAST/trauma imaging was examined in 5 studies (11.4%), while musculoskeletal applications were present in 4 studies (9.1%). Novel applications including diaphragm ultrasound for weaning prediction and airway ultrasound for cricothyrotomy landmark localization accounted for 2 studies (4.5%).

Regarding AI function across POCUS workflows, most studies investigated AI for interpretation support (34, 77.3%). Eight studies (18.2%) evaluated it for acquisition assistance, particularly for supporting less experienced operators during vascular ultrasound. Only 2 studies (4.5%) examined AI for quality assessment, such as automated scoring or image triage systems.

Fifteen studies (34.1%) deployed AI tools within active clinical workflows, including ED pathways, bedside systems, and multicenter DVT triage programs. Twenty-three studies (52.3%) did not include clinical implementation, and 6 studies (13.6%) were limited to early feasibility or proof-of-concept testing.

Regarding technology readiness, 25 studies (56.8%) were classified at TRL 4–5, corresponding to preclinical validation. Eighteen studies (40.9%) achieved TRL 6, demonstrating pro-

totype testing in live clinical settings, and 1 study (2.3%) remained at TRL 1–3, indicating early-stage algorithm development.

3.2. Diagnostic performance of AI across POCUS domains

Summary of the diagnostic performance of AI by POCUS domain is provided in table 3.

- FAST and trauma

AI systems showed high diagnostic accuracy across trauma applications. Models identified hemoperitoneum with 95% sensitivity (95% CI: 89–98%), 94% specificity (95% CI: 87–98%), and an AUC of 0.97 (95% CI: 0.94–0.99) (15). Automated detection of pericardial effusion reached 89% sensitivity (95% CI: 82–95%), 92% specificity (95% CI: 84–97%), and an AUC of 0.94 (95% CI: 0.90–0.97) (16). Pediatric FAST view classification achieved 97.8% Accuracy = 97.8% (95% CI: 96.0–99.0%). using video clips (17). An AI triage algorithm reduced expert review workload by 80% (18).

- Lung ultrasound

Across lung POCUS tasks such as B-lines, pleural effusion, lung sliding, pulmonary edema, and COVID-19 classification, AI performance was consistently high. Automated B-line detection demonstrated 95.6% sensitivity (95% CI: 92.8–98.5%) and 77.2% specificity (95% CI: 74.1–80.6%) (19). Pleural effusion detection reached sensitivities of 0.90–0.97 (95% CI range: 0.85–0.99) and specificities of 0.90–0.98 (95% CI range: 0.84–0.99) (20). Lung sliding assessment reached AUCs ranging from AUC 0.885 (95% CI: 0.828–0.956) and sensitivity of 0.921 and specificity: 0.802 (21) to 0.973 (95% CI: 0.955–0.991) (22). For pulmonary edema scoring, AI achieved similar sensitivity to physicians (95.6% vs 96.7%) but with lower specificity (23). COVID-19 classification models using X3D architectures produced F1 scores above 89% (24). Congestion scoring models correlated strongly with expert ratings ($r = 0.894$ – 0.882) (25). A transfer learning-based B-line assessment model achieved an AUC of 0.967 with 96.3% sensitivity and 92.4% specificity (64).

A landmark study by Orosz et al demonstrated that an AI ensemble achieved 100% sensitivity (95% CI: 85.8–100%) and 100% specificity (95% CI: 85.8–100%) for pneumothorax detection, significantly outperforming a panel of 11 expert clinicians in specificity (81.0% for B-mode, 73.5% for M-mode; $p = 0.02466$) (26). For pleural effusion severity, a hierarchical convolutional neural network (CNN)-geometric framework achieved 99.3% accuracy for normal vs abnormal classification and 97.0% accuracy for severity grading (27). A Hough-based structure-aware approach for B-line detection achieved F-measures of 0.837 on private and 0.808 on public datasets, outperforming YOLO and segmentation-based methods (28). For pneumothorax detection in post-chest tube patients, a CNN-based classifier demonstrated a high negative predictive value of 0.96, supporting rule-out capability (29).

- Cardiac ultrasound

AI systems achieved a sensitivity of 85.7% (95% CI: 57.2–98.2) and a specificity of 94.8% (95% CI: 90.6–97.5) for systolic dysfunction, and a sensitivity of 91.9% (95% CI: 85.6–96.0) and a specificity of 94.2% (95% CI: 87.0–98.1) for diastolic dysfunction (30). Other models attained AUROC values of 0.92 (95% CI: 0.88–0.96) for abnormal cardiac function(31), and For AI-guided detection of cardiomyopathies on POCUS, the model

achieved an AUROC of 0.903 (95% CI: 0.795–0.981) for HCM in high-confidence apical-4-chamber acquisitions in YNHHS, and an AUROC of 0.972 (95% CI: 0.959–0.983) for ATTR-CM in high-confidence parasternal short-axis acquisitions in MSHS (33).

A multi-task deep learning model (PanEcho) for echocardiography interpretation, trained on >1 million echocardiographic videos from 32,265 transthoracic echocardiographic studies (24,405 patients), achieved a median AUC of 0.91 across 18 diagnostic classification tasks and estimated LVEF with a mean absolute error of 4.2–4.5% (66). For pericardial effusion detection, UNet++ achieved 96% sensitivity, 97% specificity, 98% AUC, and 81% IoU (34). Automated LVEF measurement using a multi-frame temporal aggregation approach achieved a mean absolute error of 3.28%, comparable to inter-observer variability (35). A reinforcement learning-based view selection model matched full-study performance using 32% fewer videos for heart failure detection (67). For IVC collapse assessment, a deep learning algorithm demonstrated substantial agreement with experts ($\kappa = 0.66$ –0.78) (36). A fusion model combining heart and IVC views achieved 84.2% accuracy (AUC 0.86) for classifying diuretic need (37). In a stress test of AI generalizability, models trained on formal echocardiography underperformed on POCUS images, with an AUC of 0.74 for heart failure diagnosis (38).

- Vascular ultrasound

Across vascular assessments, novice operators guided by AI generated DVT scans with sensitivities of 95–100% (95% CI range: 89–100%) and specificities of 74–100% (95% CI range: 62–100%) (39–42). Eighty to ninety-one percent of examinations performed by novice operators obtained adequate image quality (40–42). Triage pathways supported by AI reduced the need for formal duplex ultrasound by 29–53% (39, 40). Carotid compressibility assessment reached 0.96 accuracy (95% CI: 0.88–0.99) and an F1 score of 0.97 (43).

AI-based carotid artery compressibility assessment demonstrated strong inverse correlation with systolic blood pressure ($r = -0.697$) and achieved AUCs of 0.90 for SBP <60 mmHg and 0.91 for MBP <40 mmHg, enabling noninvasive blood pressure estimation in critically ill patients (68). A domain-shift analysis highlighted substantial performance reduction when models were applied to new ultrasound systems, with kappa values falling below chance levels (-0.09 to 0.39), while human experts remained highly reliable (44).

- Musculoskeletal ultrasound

Musculoskeletal applications reported strong performance. Systems achieved accuracies of 0.889 (95% CI: 0.815–0.947) for wrist and 0.750 (95% CI: 0.652–0.838) for elbow fracture detection (45). Pediatric wrist fracture models obtained 100% sensitivity (95% CI: 88.7–100%) and approximately 0.90 specificity (95% CI: 0.82–0.96) with inference times under 2 seconds per sweep (46). A deep learning beamforming framework with bone-enhancement attention (BEAM-Net) improved contrast ratio by approximately 51% compared with single-plane-wave delay-and-sum beamforming, alongside improvements in SNR, SSI, SSIM, and EPI (47).

3.3. Novel and emerging POCUS applications

Two studies explored novel POCUS applications beyond traditional domains. Song et al. developed a multimodal AI model integrating diaphragm ultrasound video analysis with clinical indicators to predict mechanical ventilation weaning success,

achieving an AUC of 0.894, outperforming clinical indicators alone (AUC 0.746) (48). Wu et al. deployed YOLOv8n and YOLOv10n on an iPad Air achieving 13.83–14.13 FPS for real-time cricothyrotomy landmark localization in 608 patients, demonstrating the feasibility of edge AI for procedural guidance (49).

3.4. Real-time performance and feasibility

Several AI tools demonstrated real-time capability. FAST and pericardial effusion models processed frames in about 57 ms, and musculoskeletal fracture systems operated in under 2 seconds per sweep (46). Multiple studies deployed models directly on bedside hardware such as Raspberry Pi (31), WaveBase (21), or integrated ultrasound platforms.

Beyond diagnostic accuracy, AI demonstrated comparable or superior performance to human operators in several studies. Goldsmith et al. found that experts had significantly better agreement with AI than with bedside operators ($p < 0.005$) (25), while Wang et al. showed AI performed comparably to a junior sonographer (accuracy 0.872) (50).

3.5. Observed methodological limitations of included studies

Most frequently reported limitations included small sample size, restricted numbers of positive findings, single-center cohorts, and reliance on specific ultrasound vendors, all of which reduce generalizability. Many investigations analyzed stored clips instead of real-time acquisition. Technical shortcomings included difficulty handling challenging views, inaccuracies in B-line quantification, segmentation failures in certain regions, and susceptibility to image artifacts. Reference standards were often imperfect, relying on visually estimated measurements. A domain-shift analysis highlighted substantial performance reduction when models were applied to new ultrasound systems, with kappa values falling below chance levels, while human experts remained highly reliable (44).

4. Discussion

AI applications for POCUS show strong diagnostic performance across multiple indications. However, closer evaluation of the 44 studies reveals that most models were developed and tested under controlled conditions using expert-acquired images. Evidence of real-time clinical use remains limited. These findings suggest accuracy metrics may overestimate real-world performance. Several challenges remain before bedside implementation becomes practical. Notably, a landmark study demonstrated AI significantly outperforming a panel of 11 expert clinicians in pneumothorax detection (26), suggesting that human-level performance is achievable for narrowly defined tasks.

Across multiple POCUS domains, AI models frequently reported AUCs above 0.90, with particularly strong results for hemoperitoneum, pericardial effusion, pleural effusion, lung sliding, B-lines, cardiomyopathy, and pediatric fractures. Yet nearly all these results were generated from controlled settings using curated datasets, carefully selected ultrasound clips, stable imaging environments, and scans performed by expert operators. These conditions differ substantially from everyday POCUS practice. In bedside practice, image qualities vary widely due to patient factors, operator experience, and device heterogeneity. Prior work in other ultrasound fields also demonstrates how vulnerable AI is to domain shift, where performance weakens when models are exposed to unfamiliar

machines or imaging styles (51, 52). This pattern was clearly illustrated in the Blaivas (44) study within our dataset, where validated models performed with near-random accuracies when applied to other ultrasound devices. These observations suggest that current POCUS AI systems remain highly sensitive and dependent on the device, operator, and context, which raises concern about their reliability in everyday practice.

One of the most striking findings of the review is that most AI POCUS studies relied heavily on images acquired by experts or sonographers trained in fellowships. In routine practice, however, POCUS is often performed by generalists, emergency clinicians, rural practitioners, and novices, who frequently face challenges in maintaining ultrasound proficiency. Indirect evidence from a systematic review of portable ultrasound in rural and remote settings highlights operator-related variability and limitations of portable devices, but does not directly evaluate AI model performance. Similar variability has been reported in other POCUS contexts (53, 54). Despite this, many of the models with higher performance in our study were never evaluated using images generated by novice operators. Even the studies that tested AI-assisted acquisitions involved only limited pathology and small patient cohorts. This gap between the operators who generate training data and the clinicians who will ultimately rely on AI support represents a significant challenge to external validity. When training data largely come from expert-level scans, performance is unlikely to generalize well to heterogeneous images with lower quality in real clinical practice.

Live clinical implementation of AI tools in clinical POCUS practice remains uncommon among studies that we included. Only 15 of the 44 studies incorporated AI into active workflows, and these implementations were generally limited to narrow tasks such as DVT triage or small pilot evaluations. Most systems analyzed previously saved clips rather than offering actionable feedback during scanning. This pattern is consistent with broader evidence that suggests AI rarely progresses beyond retrospective validation. In the ICU literature, for example, 74% of AI studies remained at TRL $\leq$ 4, and only 2% reached TRL $\geq$ 6, with none achieving full clinical deployment (55). Similarly, our review confirms that despite strong diagnostic results, POCUS AI has yet to achieve meaningful clinical adoption, and immediate bedside integration is uncommon.

Most of the studies included in this review were small, single-center, and built on relatively uniform datasets. Positive cases were especially limited in trauma and vascular applications. While direct evidence on these limitations in POCUS AI remains limited, indirect evidence from other medical AI domains suggests that similar challenges may apply. For example, in dermatopathology, models trained on narrow or unrepresentative datasets have been shown to perform poorly on diverse populations with different skin tones (56). Likewise, prostate cancer digital twin models may fail to generalize across racial or socioeconomic groups because of biased training data (57). These examples from other domains are not direct evidence of POCUS AI performance, but they highlight general risks of dataset bias and limited generalizability that are likely relevant. POCUS AI may face analogous vulnerabilities because image appearance varies with patient body habitus, probe choice, operator skill, device manufacturer, and scanning protocols. When datasets are small and homogenous, these sources of variation are poorly represented, leading to biased models and inflated accuracy estimates.

A large proportion of models we reviewed were developed for narrow objectives, such as detecting absent lung sliding or segmenting wrist fractures. In contrast, real POCUS examinations are dynamic, with multiple views, and deeply dependent on clinical context. Single-task algorithms cannot replicate these conditions. Several studies also revealed important limitations such as inaccurate B-line counts, challenges with apical lung imaging, weakened performance with low-quality images, reliance on specific orientations, and sensitivity to artifacts. This resembles known challenges in medical AI, where the need for improved transparency in complex models remains a critical gap, particularly for AI-driven systems in cancer care (57). This instability poses a meaningful safety concern when applied in emergency care, where rapid and reliable interpretation is essential.

None of the studies in this review evaluated the safety implications of false positive or false negative outputs from AI systems. This is concerning, because prior research shows that clinicians can become overly dependent on automated recommendations, even when those tools are incorrect. A systematic review found that clinical decision support tools can trigger automation bias, causing users to abandon correct judgments in favor of incorrect recommendations (58). Further work has shown that automation bias can occur even in simple decision tasks that require complex verification (59), and that poor alignment between trust and confidence plays a major role in this regard (60). Considering that POCUS informs urgent decisions such as ruling out pneumothorax, the absence of any analysis on automation bias is a substantial gap in the current evidence base.

4.1. Limitations

This review has several limitations that should be considered when interpreting our findings. First, despite comprehensive searching of five major databases, we may have missed relevant studies due to our exclusion of grey literature (conference abstracts, dissertations, technical reports), restriction to English-language publications, and exclusion of pre-2020 studies. While the 2020 cutoff was justified by the rapid recent advances in AI, it may have excluded important foundational work.

Second, the field has a strong publication bias toward positive results. Studies reporting negative or null findings are less likely to be published, which may inflate the apparent success of AI applications in POCUS. This was evident in our analysis, where nearly all studies reported high diagnostic performance, creating potential for overestimation of real-world effectiveness.

Third, the heterogeneity in outcome reporting across studies (different performance metrics, variable reference standards, diverse POCUS applications) made it challenging to compare results directly. This heterogeneity is inherent to the field and justifies our scoping review approach rather than meta-analysis.

Fourth, the lack of standardized terminology for AI roles (e.g., "decision support," "interpretation assistant," "second reader" are used interchangeably) may have resulted in some misclassification of studies, despite our best efforts to apply consistent criteria.

Fifth, our search ended on August 17, 2026, and new studies published since may provide additional evidence not captured in this review.

sixth, we did not conduct formal risk-of-bias assessments using established tools (e.g., Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2, the Prediction model Risk Of Bias ASsessment Tool (PROBAST)) as these are designed for diagnostic accuracy systematic reviews and are not fully appropriate for scoping reviews.

Finally, some of the limitations discussed in this review are drawn from indirect evidence in other medical AI domains, as direct POCUS-specific data on issues such as dataset bias, domain shift, and automation bias remain sparse."

4.2. Future directions

Although performance metrics are impressive, the overall evidence points to a considerable distance between current AI POCUS capability and true clinical readiness. Future research should prioritize multicenter and multi-device validation to address vulnerabilities with domain shift. Larger, more diverse datasets are also essential to reduce bias and improve generalizability. Since many AI POCUS tools are intended for less experienced operators, studies should include scans acquired by novice operators, and live evaluations at bedside rather than retrospective analysis. Finally, integrating safety testing and automation bias assessment is critical before widespread deployment.

5. Conclusions

This review demonstrates that AI POCUS achieves high diagnostic performance across multiple applications but remains limited by small datasets, operator dependence, and poor generalizability. These findings resonate with broader challenges in medical AI, where most systems stall before clinical integration. Strengthening dataset diversity, external validation, and safety evaluation is necessary for AI POCUS to deliver reliable, practical clinical value.

6. Declarations

6.1. Acknowledgments

Not applicable.

6.2. Conflict of interests

The authors declare that they have no competing interests.

6.3. Funding and supports

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

6.4. Authors' contributions

Milad Bakhshi: Conceptualization, methodology, formal analysis, investigation, data curation, writing - original draft, visualization Alireza Majidi: Conceptualization, methodology, validation, investigation, writing - review & editing Saba Aghajani: Methodology, validation, investigation, data curation, writing - review & editing All authors read and approved the final version of the manuscript.

6.5. Ethical statement

Not applicable.

6.6. Informed consent

None.

6.7. Availability of supporting data

All data generated or analyzed during this scoping review are included in this published article and its supplementary information files.

6.8. Using artificial intelligence chatbots statement

The authors used Grammarly for grammar and language editing assistance during the preparation of this manuscript. This tool was employed exclusively for proofreading and did not influence the scientific content, interpretation of findings, or decision-making. No generative AI tools (including ChatGPT, Claude, Gemini, or similar) were used for drafting, data analysis, literature synthesis, or writing.

7. References

1. Chang H, Ko E, Lee JH, Kim M, Kim T, Shin TG, et al. Emergency department crowding: a national data report. *Clin Exp Emerg Med*. 2024;11(4):331-334. doi:10.15441/ceem.24.337.
2. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS One*. 2018;13(8):e0203316. doi:10.1371/journal.pone.0203316
3. Alfoti BOO, Alfoti FOO, Alothman STH, Al-Dhafiri TMA, Al-Harbi NHM, Al-Khalidi AM, et al. Utilization of Point-of-Care Ultrasound (POCUS) in Emergency and Critical Care: Role of Nursing for Enhancing Diagnostic Accuracy and Efficiency-Systematic Review. *Egyptian Journal of Chemistry* 2024;67(13): 705-16.doi: 10.21608/ejchem.2024.297100.9851
4. Su J, Tie X, Wei Y, Zhou R, Zou T, Qin Y, et al. Critical care ultrasound: development, evolution, current and evolving clinical concepts in critical care medicine. *Frontiers in medicine*. 2025; 12:1622604.
5. Platt A, Peterson T, Peethumnongsin E, Kuchibhatla M, Theophanous RG. Prospective cohort study identifies barriers and facilitators to point-of-care ultrasound use in an academic emergency department. *BMJ Open Qual*. 2026;15(3):e004210. doi:10.1136/bmjoq-2026-004210.
6. Gao Y, Zhang H, Wu C, Zhang Z, Pan L, He X, et al. Point of care ultrasound knowledge attitudes and practices among emergency and ICU healthcare providers. *Sci Rep*. 2025;15(1):33378.
7. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA, et al. A survey of deep-learning applications in ultrasound: Artificial intelligence-powered ultrasound for improving clinical workflow. *J Am Coll Radiol*. 2019;16(9 Pt B):1318-1328. doi:10.1016/j.jacr.2019.06.004
8. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):e271-e297. doi:10.1016/S2589-7500(19)30123-2.
9. Wistrom, R., Khait, L. & Nelson, G. Artificial Intelligence in Point-of-care Ultrasound. *Curr Emerg Hosp Med Rep* 12, 89-94 (2024). <https://doi.org/10.1007/s40138-024-00293-8>
10. Toporek GA, Wang H, inventors. US Patent 11,712,220 B2. 2023.
11. Kayarian F, Patel D, O'Brien JR, Schraft EK, Gottlieb M. Artificial intelligence and point-of-care ultrasound: Benefits, limitations, and implications for the future. *Am J Emerg Med*. 2024;80:119-22.
12. Kim S, Fischetti C, Guy M, Hsu E, Fox J, Young SD. Artificial Intelligence (AI) Applications for Point of Care Ultrasound

- (POCUS) in Low-Resource Settings: A Scoping Review. *Diagnosics*. 2024;14(15):1669.
13. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-73.
 14. Mankins JC. Technology readiness levels: a white paper. Washington, DC: NASA; 1995.
 15. Leo MM, Potter IY, Zahiri M, 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(5):2035-50. doi: 10.1007/s10278-023-00845-6
 16. 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(8):1947-1959. doi: 10.1007/s11517-023-02855-6
 17. Kornblith AE, Addo N, Dong R, Rogers R, Grupp-Phelan J, Butte A, et al. Development and Validation of a Deep Learning Strategy for Automated View Classification of Pediatric FAST. *J Ultrasound Med*. 2022;41 (8):1915-24.
 18. Cull J, Morrow D, Manasco C, Vaughan A, Eicken J, Smith H, et al. A quality assessment tool for focused abdominal sonography for trauma examinations using artificial intelligence. *J Trauma Acute Care Surg*. 2025;98(1):111-6.
 19. Ienghong K, Cheung LW, Gaysonsiri D, Apiratwarakul K. The diagnostic performance of automatic B-lines detection for evaluating pulmonary edema in the emergency department among novice point-of-care ultrasound practitioners. *Emerg Radiol*. 2025;32(2):241-246. doi:10.1007/s10140-025-02319-4.
 20. Chaudhary R, Ho J, Smith D, Hossain S, Hargun J, VanBerlo B, et al. Diagnostic accuracy of an automated classifier for the detection of pleural effusions in patients undergoing lung ultrasound. *AmJEmergMed*. 2025;90:142-150. doi:10.1016/j.ajem.2025.01.041.
 21. Clausdorff Fiedler H, Prager R, Smith D, Wu D, Dave C, Tschirhart J, et al. Automated real-time detection of lung sliding using artificial intelligence: a prospective diagnostic accuracy study. *Chest*. 2024;166(2):362-370. doi:10.1016/j.chest.2024.02.011.
 22. VanBerlo B, Derek Wu, Brian Li, Marwan A Rahman, Gregory Hogg, VanBerlo, ., et al. Accurate assessment of the lung sliding artefact on lung ultrasonography using a deep learning approach. *Comput Biol Med*. 2022;148:105953. doi:10.1016/j.compbiomed.2022.105953.
 23. Gottlieb M, Patel D, Viars M, Tsintolas J, Peksa GD, Bailitz J. Comparison of artificial intelligence versus real-time physician assessment of pulmonary edema with lung ultrasound. *Am J EmergMed*. 2023;70:109-112. doi:10.1016/j.ajem.2023.05.029.
 24. Gazzoni M, La Salvia M, Torti E, Secco G, Perlini S, Leporati F. Perceptive SARS-CoV-2 End-To-End Ultrasound Video Classification through X3D and Key-Frames Selection. *Bioengineering(Basel)*. 2023;10(3):282. doi:10.3390/bioengineering10030282
 25. Goldsmith AJ, Jin M, Lucassen R, Duggan NM, Harrison NE, Wells W, et al. Comparison of pulmonary congestion severity using artificial intelligence-assisted scoring versus clinical experts: A secondary analysis of BLUSHED-AHF. *Eur J Heart Fail*. 2023;25(7):1166-1169. doi:10.1002/ehf.2881.
 26. Orosz G, Szabó RZ, Szabó M, Gyombolai P, Tóth JT, Ruttkay T, et al. Explainable transfer learning ensemble AI model for lung ultrasound pneumothorax detection with expert benchmark. *Scand J Trauma Resusc Emerg Med*. 2026;34(1):95. doi:10.1186/s13049-026-01614-4.
 27. Lin YC, Fang YH, Chen TY, Wang JT, Lin CM, Chen YS, et al. A Hierarchical CNN-Geometric Framework for Pleural Effusion Severity Classification in Lung Ultrasound. *IEEE Sens J*. 2026;1-1. doi:10.1109/JSEN.2026.3721141.
 28. Liu T, Zhou H, Wang JH, Zhang Y, Chen C, Chen Y, et al. HSD: Hough-Based Structure-Aware Detection of B-Lines in Lung Ultrasound. *IEEE J Biomed Health Inform*. 2026; Epub ahead of print. doi:10.1109/JBHI.2026.3684930.
 29. Cote M, Smith D, Orozco N, Huggard B, Wu B, Tran K, et al. Diagnostic Accuracy of Automated Pneumothorax Detection Via Novice-Acquired Ultrasound After Chest Tube Removal: Comparison With Expert Interpretation and Chest Radiography. *CHEST Pulmonary*. 2026;4(2):100246. doi: 10.1016/j.chpulm.2026.100246
 30. Gottlieb M, Schraft E, O'Brien J, Patel D. Diagnostic accuracy of artificial intelligence for identifying systolic and diastolic cardiac dysfunction in the emergency department. *Am J Emerg Med*. 2024;86:115-119. doi:10.1016/j.ajem.2024.10.019.
 31. He B, Dash D, Duanmu Y, Tan TX, Ouyang D, Zou J. AI-enabled assessment of cardiac function and video quality in emergency department point-of-care echocardiograms. *J EmergMed*. 2024;66(2):184-191. doi:10.1016/j.jemermed.2023.02.005.
 32. Oikonomou EK, Vaid A, Holste G, Coppi A, McNamara RL, Baloesu C, et al. Artificial intelligence-guided detection of under-recognised cardiomyopathies on point-of-care cardiac ultrasonography: a multicentre study. *Lancet Digit Health*. 2025;7(2):e113-e123. doi:10.1016/S2589-7500(24)00249-8.
 33. Lin X, Yang F, Chen Y, Chen X, Wang W, Chen X, et al. Echocardiography-based AI detection of regional wall motion abnormalities and quantification of cardiac function in myocardial infarction. *Front Cardiovasc Med*. 2022;9:903660. doi:10.3389/fcvm.2022.903660
 34. Chou HH, Chang YC, Lien WC, Lin LC, Lin XZ, Hsu TE, et al. Efficient Deep Learning Models Revolutionize Doctor's Training for Point-of-Care Ultrasound. *IEEE Access*. 2025;13:76038-76046. doi:10.1109/ACCESS.2025.3562674.
 35. Shan C, Guo G, Shi Z, Qi Z, Chen F, Cheng H, et al. ESPNet: an automated left ventricular ejection fraction measurement via multi-frame temporal aggregation. *IEEE J Biomed Health Inform*. 2026; In press. doi:10.1109/JBHI.2026.3715789.
 36. Blaivas M, Blaivas LN, Tsung JW. Deep learning algorithm performance compared to experts in visual evaluation of inferior vena cava collapse on ultrasound to determine intravenous fluid need in dehydration management. *Signa Vitae*. 2021;17(5):34.
 37. Wshah S, Xu B, Steinharter J, Reilly C, Morrisette K. Classification of clinically relevant intravascular volume status using point of care ultrasound and machine learning. *J Med Imaging (Bellingham)*. 2022;9(5):054502.
 38. Crockett D, Kelly C, Brundage J, Jones J, Ockerse P. A stress test of artificial intelligence: can deep learning models trained from formal echocardiography accurately interpret point-of-care ultrasound? *J Ultrasound Med*. 2022;41(12):3003-3012. doi:10.1002/jum.16003.
 39. Avgerinos E, Spiliopoulos S, Psachoulia F, Yfantis A, Plakas G, Grigoriadis S, et al. Novel artificial intelligence guided non-expert compression ultrasound deep vein thrombosis diagnostic pathway may reduce vascular laboratory venous testing. *Eur J Vasc Endovasc Surg*. 2025;70(4):517-522. doi:10.1016/j.ejvs.2025.04.070.

40. Oppenheimer J, Mandegaran R, Staabs F, Adler A, Singöhl S, Kainz B, et al. Remote expert DVT triaging of novice-user compression sonography with AI-guidance. *Ann Vasc Surg*. 2024;99:272-279. doi:10.1016/j.avsg.2023.08.022.
41. Speranza G, Mischkewitz S, Al-Noor F, Kainz B. Value of clinical review for AI-guided deep vein thrombosis diagnosis with ultrasound imaging by non-expert operators. *NPJ Digit Med*. 2025;8(1):135. doi:10.1038/s41746-025-01518-0.
42. Nothnagel K, Aslam MF. Evaluating the benefits of machine learning for diagnosing deep vein thrombosis compared with gold standard ultrasound: a feasibility study. *BJGP Open*. 2025;8(4): BJGPO.2024.0057. doi: 10.3399/BJGPO.2024.0057.
43. Park, S., Yoon, H., Yeon Kang, S., Joon Jo, I., Heo, S., Chang, H., Eun Park, J et al. Artificial intelligence-based evaluation of carotid artery compressibility via point-of-care ultrasound in determining the return of spontaneous circulation during cardiopulmonary resuscitation. *Resuscitation*. 2024;202:110302. doi:10.1016/j.resuscitation.2024.110302.
44. Blaivas M, Blaivas LN, Tsung JW. Deep Learning Pitfall: Impact of Novel Ultrasound Equipment Introduction on Algorithm Performance and the Realities of Domain Adaptation. *J Ultrasound Med*. 2022;41(4):855-863. doi:10.1002/jum.15765.
45. Zhou Y, Knight J, Alves-Pereira F, Keen C, Hareendranathan AR, Jaremko JL. Wrist and elbow fracture detection and segmentation by artificial intelligence using point-of-care ultrasound. *J Ultrasound*. 2025;28(2):465-473. Doi:10.1007/s40477-025-01019-6.
46. Zhang J, Boora N, Melendez S, Rakkunedeth Hareendranathan A, Jaremko J. Diagnostic Accuracy of 3D Ultrasound and Artificial Intelligence for Detection of Pediatric Wrist Injuries. *Children (Basel)*. 2021;8(6):431. doi:10.3390/children8060431.
47. Madhusoodanan M, Panicker MR, Gopalakrishnan PH, Hareendranathan AR. Beam-net: a deep learning framework with bone enhancement attention mechanism for high resolution high frame rate ultrasound beamforming. *arXiv preprint arXiv:2507.15306*. 2025. doi:10.48550/arXiv.2507.15306.
48. Song F, Liu H, Ma H, Chen X, Wang S, Qin T, et al. AI model based on diaphragm ultrasound to improve the predictive performance of invasive mechanical ventilation weaning: prospective cohort study. *JMIR Form Res*. 2025;9:e72482. doi: 10.2196/72482
49. Wu C-Y, Li J-D, Shih P-Y, Huang C-C, Cheng H-L, Wu C-Y, et al. Real-Time Identification of Cricothyrotomy Landmarks in Emergency Care and Obstetric Patients Using Wireless Handheld Ultrasound and Edge-Computing Artificial Intelligence: A Prospective Observational Study. *J Med Syst*. 2025;49(1):131.
50. Wang Y, Ye R, Cai Q, Wu Y, Gong L, Li Z, et al. Deep Learning for Ultrasound-Based Auxiliary Diagnosis of Emergency Ascites. *Ultrasound Med Biol*. 2026;52(4):782-8.
51. Komatsu M, Teraya N, Natsume T, Harada N, Takeda K, Hamamoto R. Clinical Application of Artificial Intelligence in Ultrasound Imaging for Oncology. *Jma j*. 2025;8(1):18-25.
52. Zeng H, Zou K, Chen Z, Gao Y, Chen H, Zhang H, et al. Training-Free Image Style Alignment for Domain Shift on Handheld Ultrasound Devices. *IEEE Trans Med Imaging*. 2025;44(4):1942-52.
53. Shaddock L, Smith T. Potential for Use of Portable Ultrasound Devices in Rural and Remote Settings in Australia and Other Developed Countries: A Systematic Review. *J Multidiscip Healthc*. 2022;15:605-25.
54. Sorensen B, Hunskaar S. Point-of-care ultrasound in primary care: a systematic review of generalist performed point-of-care ultrasound in unselected populations. *Ultrasound J*. 2019;11(1):31.
55. Berkhout WEM, van Wijngaarden JJ, Workum JD, van de Sande D, Hilling DE, Jung C, et al. Operationalization of Artificial Intelligence Applications in the Intensive Care Unit: A Systematic Review. *JAMA Netw Open*. 2025;8(7):e2522866.
56. Mehta A, Motavaf M, Raza D, Jairath N, Pulavarty A, Xu Z, et al. Deep Learning Image Processing Models in Dermatopathology. *Diagnostics (Basel)*. 2025;15(19):2517. doi:10.3390/diagnostics15192517.
57. John A, Alhaji R, Rokne J. A systematic review of AI as a digital twin for prostate cancer care. *Comput Methods Programs Biomed*. 2025;268:108804.
58. Goddard K, Roudsari A, Wyatt J. Decision support and automation bias: methodology and preliminary results of a systematic review. *Stud Health Technol Inform*. 2011;164:3-7.
59. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association: JAMIA*. 2017;24(2):423-31.
60. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association: JAMIA*. 2012;19(1):121-7.
61. Thyagachandran A, Lefchak B, Murthy HA, Bergmann KR, Madhok M. Development of a deep learning model for intussusception using point-of-care ultrasound. *Frontiers in radiology*. 2026;6:1812587.
62. Marini TJ, Nayaruparambil AA, Baran TM, Murali T, Drury S, Zhao YT, et al. New lung ultrasound system for rapid triage of pulmonary disease without a radiologist or sonographer. *BMC Pulm Med*. 2025;25(1):499.
63. Russell FM, Ehrman RR, Barton A, Sarmiento E, Ottenhoff JE, Nti BK. B-line quantification: comparing learners novice to lung ultrasound assisted by machine artificial intelligence technology to expert review. *Ultrasound J*. 2021;13(1):33.
64. Pare JR, Gjestebj LA, Tonelli M, Leo MM, Muruganandan KM, Choudhary G, et al. Transfer Learning-Based B-Line Assessment of Lung Ultrasound for Acute Heart Failure. *Ultrasound Med Biol*. 2024;50(6):825-832. doi: 10.1016/j.ultrasmedbio.2024.02.004
65. Wang E, Wu J, Gao S, Zhou P, Chen Y, Zhou G. SM-FSL: Similarity Guided Multi-Source Few-Shot Learning for Lung Ultrasound Diagnosis. *IEEE J Biomed Health Inform*. 2026; Epub ahead of print. doi:10.1109/JBHI.2026.3721499.
66. Holste G, Oikonomou EK, Tokodi M, Kovács A, Wang Z, Khara R. Complete AI-Enabled Echocardiography Interpretation With Multitask Deep Learning. *JAMA*. 2025;334(4):306-318. doi:10.1001/jama.2025.8731.
67. Saadat A, Hashemi N, Khodabakhshian B, Tsang MY, Luong C, Tsang TSM, et al. Optimizing point-of-care ultrasound video acquisition for probabilistic multi-task heart failure detection. *Int J Comput Assist Radiol Surg*. 2026;21(5):1151-1159. doi:10.1007/s11548-026-03680-6.
68. Maeng SJ, Park S, Jo IJ, Lee G, Hwang SY, Chung MJ, et al. AI-driven carotid artery compressibility assessment via point-of-care ultrasound for blood pressure estimation in critically ill and post-resuscitation patients: a prospective observational study. *Crit Care*. 2026;30:2. doi:10.1186/s13054-025-05772-2

Table 1: Population-Concept-Context (PCC) framework and main search strategy

Search Block	PCC Component	Specific Terms & MeSH	Rationale for Inclusion
Artificial Intelligence & Machine Learning	Concept	MeSH Terms: "Artificial Intelligence"[Mesh], "Deep Learning"[Mesh], "Neural Networks, Computer"[Mesh], "Machine Learning"[Mesh] Keywords: "convolutional neural network"[tiab], "CNN"[tiab], "deep learning"[tiab], "neural network"[tiab], "computer vision"[tiab], "AI"[tiab]	Captures the core technological intervention. Using both broad MeSH terms and specific technical keywords ensures comprehensive coverage of the relevant AI methods.
Point-of-Care Ultrasound (POCUS)	Population & Concept	MeSH Terms: "Ultrasonography"[Mesh], "Point-of-Care Systems"[Mesh], "Echocardiography"[Mesh] Keywords: "POCUS"[tiab], "point-of-care ultrasound"[tiab], "bedside ultrasound"[tiab], "focused ultrasound"[tiab], "E-FAST"[tiab], "FAST exam"[tiab], "Focused Assessment with Sonography for Trauma"[tiab], "lung ultrasound"[tiab]	Defines the medical domain and the specific type of examination. Keywords include acronyms and common phrases to capture all relevant POCUS literature.
Emergency Department & Triage	Context	MeSH Terms: "Emergency Service, Hospital"[Mesh], "Triage"[Mesh] Keywords: "emergency department"[tiab], "ED"[tiab], "triage"[tiab], "emergency room"[tiab], "ER"[tiab], "resuscitation"[tiab]	Specifies the high-acuity, time-pressured clinical setting. This filter ensures results are relevant to the immediate assessment and triage phase.

The table outlines the search blocks, corresponding PCC components, specific Medical Subject Headings (MeSH) and keywords used, and the rationale for their inclusion in the scoping review.

Table 2: Summary of included studies investigating AI/ML as a second reader for focused POCUS in emergency department triage. The table presents the study reference, POCUS application, primary AI role, sample size, key AI performance metrics, whether clinical implementation was tested, and the assessed Technology Readiness Level (TRL)

Study (Author, Year)	POCUS Application	AI's Primary Role	Sample Size (Patients / Clips)	Key AI Performance Metrics	Clinical Implementation?	TRL
FAST and Trauma						
Cull & Morrow, 2025 (18)	FAST	Quality Assessor	1,514 clips	Reduced expert review need by 80%	No (Under development)	6
Leo et al., 2023 (15)	FAST	Interpretation Assistant	94 patients	Sens: 95%, Spec 94%, AUC: 0.97	No	4-5
Kornblith et al., 2022 (17)	Pediatric FAST	Acquisition Guide	699 studies	View classification accuracy: 97.8%	No	4-5
Thyagachandran et al., 2026 (61)	Abdominal (Intussusception)	Interpretation Assistant	49 patients	[Model A]: Sens 88%; [Model B]: Accuracy 90%	No	4-5
Wang et al., 2026 (50)	FAST (Ascites)	Interpretation Assistant	848 external-validation images (424 positive, 424 negative)	Acc: 0.883, Sens: 0.901, Spec: 0.837, AUC: 0.871	No	4-5
Lung Ultrasound						
Marini et al., 2025 (62)	Lung	Acquisition Guide	119 clips (67 patients) total; performance from 36 test clips (20 subjects)	Acc: 91.7%, Sens: 85.7%, Spec 95.5%	No	4-5
Ienghong et al., 2025 (19)	Lung (B-lines)	Interpretation Assistant	66 patients	Sens :95.6%, Spec: 77.2%	Yes (Retrospective)	4
Chaudhary et al., 2025 (20)	Lung (Pleural effusion)	Interpretation Assistant	785 patients	Sens :0.90-0.97, Spec :0.90-0.98	No	4-5
Clausdorff Fiedler et al., 2024 (21)	Lung (Lung sliding)	Interpretation Assistant	62 patients	Sens: 0.921, Spec :0.802, AUC: 0.885	Yes (Bedside hardware)	6
Goldsmith et al., 2023 (25)	Lung (Congestion scoring)	Interpretation Assistant	130 patients	r=0.882-0.894 with experts	No	4-5
Russell et al., 2021 (63)	Lung (B-lines)	Interpretation Assistant	29 patients 3,075 clips from 738 patients (training: 2,535 clips from 614 patients; test: 540 clips from 124 patients)	ICC:0.56 with expert	No	4-5
VanBerlo et al., 2022 (22)	Lung (Lung sliding)	Interpretation Assistant		Sens: 93.5%, Spec: 87.3%, AUC :0.973	No	4-5
Pare et al., 2024 (64)	Lung (B-lines)	Interpretation Assistant	60 patients	AUC :0.967, Sens :96.3%, Spec: 92.4%	No	4-5
Gazzoni et al., 2023 (24)	Lung (COVID-19)	Interpretation Assistant	450 patients	F1-Score :>89%	No	4-5
Gottlieb et al., 2023 (23)	Lung (Pulmonary edema)	Interpretation Assistant	71 patients	Sens: 95.6% (AI) vs 96.7% (MD)	No	4-5

Study (Author, Year)	POCUS Application	AI's Primary Role	Sample Size (Patients / Clips)	Key AI Performance Metrics	Clinical Implementation?	TRL
Cote et al., 2026 (29)	Lung (Pneumothorax)	Interpretation Assistant	76 patients (64 validation)	Sens: 0.775, Spec: 0.831, NPV 0.96	No	4-5
Orosz et al., 2026 (26)	Lung (Pneumothorax)	Interpretation Assistant	1,856 M-mode images	Sens: 100%, Spec: 100% (vs experts 95.8%/81.0%)	Yes (Real-time version)	6
Lin et al., 2026 (27)	Lung (Pleural effusion severity)	Interpretation Assistant	365 images	Stage 1 Acc: 99.3%, Stage 2 Acc :97.0%	No	4-5
Wang et al., 2026 (65)	Lung (COVID-19)	Interpretation Assistant	1,486 images	Accuracy :78.04% (few-shot)	No	4-5
Liu et al., 2026 (28)	Lung (B-lines)	Interpretation Assistant	973 images	F-measure: 0.837 (private), 0.808 (public)	No	4-5
Cardiac Ultrasound						
Holste et al., 2025 (66)	Cardiac	Interpretation Assistant	1.2 million echocardiographic videos from 32,265 TTE studies of 24,405 patients	Median AUC: 0.91 for 18 tasks	Yes (Retrospective validation)	6
Gottlieb et al., 2024 (30)	Cardiac	Interpretation Assistant	209 patients	Systolic: Sens: 85.7%, Spec: 94.8%; Diastolic: Sens 91.9%, Spec 94.2%	Yes (ED workflow)	5
Oikonomou et al., 2025 (32)	Cardiac	Interpretation Assistant	~39k patients	AUROC HCM: 0.903; ATTR-CM: 0.972	Yes (Retrospective validation)	6
Potter et al., 2023 (16)	Cardiac (Pericardial effusion)	Interpretation Assistant	76 patients	Sens: 89%, Spec :92%, AUC :0.94	No	4-5
He et al., 2024 (31)	Cardiac	Interpretation Assistant	573 patients	AUROC: 0.92 (abnormal function); 0.96 in pilot	Yes (Prospective pilot)	6
Crockett et al., 2022 (38)	Cardiac	Interpretation Assistant	306 patients	AUC :0.74 (HF diagnosis); Dice: 0.72	No	4-5
Lin et al., 2022 (33)	Cardiac	Interpretation Assistant	6,953 exams	RWMA AUC 0.90-0.91; LVEF bias ~4%	No	6
Shan et al., 2026 (35)	Cardiac (LVEF)	Interpretation Assistant	250 patients	MAE :3.28%, Corr: 0.982	No	4-5
Chou et al., 2025(34)	Cardiac (Pericardial effusion)	Interpretation / Quality Assessor	101 patients	Sens :96%, Spec :97%, AUC: 98%, IoU: 81%	Yes (Educational)	6
Saadat et al., 2026 (67)	Cardiac (View selection)	Acquisition Guide	12,180 patient-level studies	Matches full performance with 32% fewer videos	No	4-5
Blaivas et al., 2021 (36)	Cardiac (IVC)	Interpretation Assistant	184 videos	Agreement κ =0.66-0.78 with experts	Yes (Real-time algorithm)	6
Vascular Ultrasound						
Avgerinos et al., 2025 (39)	Vascular (DVT)	Acquisition Guide	53 patients	Sens :100%, Spec: 95.7%; 32% avoided DUS	Yes (ED pathway)	6
Speranza et al., 2025 (41)	Vascular (DVT)	Acquisition Guide	381 patients	EM physicians: Sens 95-98%, Spec: 97-100%	Yes (Multicenter UK)	6
Oppenheimer et al., 2024 (40)	Vascular (DVT)	Acquisition Guide	73 patients	Sens :100%, Spec: 95.1%; 53% avoided formal scan	Yes (2 hospital sites)	6
Nothnagel et al., 2025 (42)	Vascular (DVT)	Acquisition Guide	91 patients	Sens: 100%, Spec :91%; 53% low risk	Yes (Feasibility)	6
Park et al., 2024 (43)	Vascular (Carotid)	Interpretation Assistant	54 patients	Acc: 0.96, F1-score :0.97	No	5
Blaivas et al., 2022 (44)	Vascular (IVC)	Quality Assessor	42 new videos	Kappa fell to -0.09-0.39 on new devices	No	1-3
Maeng et al., 2026 (68)	Vascular (Carotid/BP)	Interpretation Assistant	56 patients	AUC :0.90 (SBP<60); RMSE: 8.3 mmHg	No	4-5
Musculoskeletal Ultrasound						
Zhou et al., 2025 (45)	MSK (Wrist/Elbow)	Interpretation Assistant	55 patients	Wrist Acc: 0.889; Elbow Acc: 0.750	Yes (Pediatric ED)	4
Zhang et al., 2021 (46)	MSK (Pediatric wrist)	Interpretation Assistant	30 patients	Sens :1.00, Spec ~0.90	No	4-5
Madhusoodanan et al., 2025 (47)	MSK (Bone enhancement/beamforming)	Image reconstruction/enhancement	1,900 in-vivo wrist datasets; 1,200 synthetic datasets	CR improved ~51% vs SPW-DASB; SNR, SSI, SSIM, EPI improved; no radiologist rating reported	No	4-5
Novel Applications						
Wu et al., 2025 (49)	Airway (Cricothyrotomy)	Acquisition Guide	608 patients	AUC :>0.88,14 FPS on iPad	Yes (Edge deployment)	6
Song et al., 2025 (48)	Diaphragm (Weaning)	Interpretation Assistant	88 patients	AUC :0.894 (multimodal)	No	4-5
Other Applications						

Study (Author, Year)	POCUS Application	AI's Primary Role	Sample Size (Patients / Clips)	Key AI Performance Metrics	Clinical Implementation?	TRL
Wshah et al., 2022 (37)	Cardiac/Thoracic/IVC	Interpretation Assistant	101 patients	Fusion Acc :84.2%, AUC :0.86	No	4-5
AI, artificial intelligence; ML, machine learning; POCUS, point-of-care ultrasound; FAST, focused assessment with sonography for trauma; ED, emergency department; TRL, technology readiness level; Sens, sensitivity; Spec, specificity; AUC, area under the receiver operating characteristic curve; Acc, accuracy; NPV, negative predictive value; ICC, intraclass correlation coefficient; MAE, mean absolute error; Corr, correlation; RWMA, regional wall motion abnormality; LVEF, left ventricular ejection fraction; HCM, hypertrophic cardiomyopathy; ATTR-CM, transthyretin amyloid cardiomyopathy; DVT, deep vein thrombosis; IVC, inferior vena cava; MSK, musculoskeletal; BP, blood pressure; SBP, systolic blood pressure; MBP, mean blood pressure; RMSE, root mean square error; FPS, frames per second; IoU, intersection over union; CR, contrast ratio; HF, heart failure; COVID-19, coronavirus disease 2019; EM, emergency medicine; UK, United Kingdom; MD, medical doctor; CI, confidence interval; κ, Cohen's kappa; r, correlation coefficient; vs, versus.						

Table 3: Summary of diagnostic performance by POCUS (Point-of-care Ultrasound) domain. The table aggregates key performance metrics (Sensitivity (sens), Specificity (spec), area under the receiver operating characteristic (ROC) curve (AUC)) for primary AI-assisted tasks across major POCUS application areas, with representative citations

Citations	Primary Task	Key Performance Metrics
FAST & Trauma		
(15)	Hemoperitoneum detection	Sens: 95%, Spec: 94%, AUC: 0.97
(16)	Pericardial effusion detection	Sens: 89%, Spec: 92%, AUC: 0.94
(17)	View classification (Pediatric FAST)	Accuracy: 97.8%
(18)	Quality triage	Reduced expert review by 80%
(50)	Ascites / Free intraperitoneal fluid	External Acc: 0.883, AUC: 0.871
(61)	Intussusception (target sign)	[Model A] Sens 88%; [Model B] Accuracy 90%
Lung Ultrasound		
(19)	B-line detection	Sens: 95.6% (95% CI: 92.8–98.5); Spec: 77.2% (95% CI: 74.1–80.6)
(28)	B-line detection (Hough-based)	F-measure: 0.837 (private), 0.808 (public)
(20)	Pleural effusion detection	Pleural effusion detection — General model: Sens 0.90, Spec 0.89; Large effusion model: Sens 0.97, Spec 0.90; Trauma model: Sens 0.91–0.97
(27)	Pleural effusion severity classification	Stage 1 Acc: 99.3%; Stage 2 Acc: 97.0%
(21, 22)	Lung sliding assessment	AUC: 0.885 – 0.973
(26)	Pneumothorax detection	Sens: 100%, Spec: 100%
(29)	Pneumothorax detection (post-chest tube)	Sens: 0.775, Spec: 0.831, NPV: 0.96
(23)	Pulmonary edema scoring	Sens: 95.6% (AI) vs 96.7% (MD); Spec: 64.1% vs 79.1%
(24)	COVID-19 classification	F1-Score > 89%
(65)	COVID-19 diagnosis (few-shot learning)	Accuracy: 78.04% (3-way 5-shot)
(25)	Congestion scoring	Congestion scoring: $r = 0.894$ (AI vs E1), $r = 0.882$ (AI vs E2)
(62)	Lung triage (Normal/Abnormal)	Acc: 91.7%, Sens: 85.7%, Spec: 95.5%
(64)	B-line assessment (Transfer learning)	AUC: 0.967, Sens: 96.3%, Spec: 92.4%
Cardiac Ultrasound		
(30)	Systolic dysfunction detection	Cardiac dysfunction: Systolic — Sens 85.7% (95% CI 57.2–98.2), Spec 94.8% (95% CI 90.6–97.5); Diastolic — Sens 91.9% (95% CI 85.6–96.0), Spec 94.2% (95% CI 87.0–98.1)
(30)	Diastolic dysfunction detection	Sens: 91.9%, Spec: 94.2%
(31)	Abnormal cardiac function detection	AUROC: 0.92 (0.96 in prospective pilot)
(32)	HCM detection	AUROC: up to 0.903
(32)	ATTR-CM detection	HCM detection (high-confidence A4c, YNHHS): AUROC 0.903 (95% CI: 0.795–0.981); MSHS: 0.890 (95% CI: 0.839–0.938) ATTR-CM detection (high-confidence PSX, MSHS): AUROC 0.972 (95% CI: 0.959–0.983); YNHHS: 0.907 (95% CI: 0.874–0.932)
(33)	Regional wall motion abnormality	AUC: 0.90–0.91; LVEF bias ~4%
(34)	Pericardial effusion detection (UNet++)	Sens: 96%, Spec: 97%, AUC: 98%, IoU: 81%
(35)	LVEF measurement (multi-frame)	MAE: 3.28%, Corr: 0.982
(67)	AI-guided view selection	Full performance using 32% fewer videos
(66)	Multitask echocardiography interpretation	Median AUC: 0.91 for 18 tasks; LVEF MAE: 4.2–4.5%
(38)	POCUS stress test (TTE-trained model)	AUC: 0.74 (HF diagnosis); Dice: 0.72
(36)	IVC collapse assessment	Agreement $\kappa=0.66$ –0.78 with experts
(37)	Diuretic need classification (Heart+IVC)	Acc: 84.2%, AUC: 0.86
Vascular Ultrasound		
(39–42)	DVT diagnosis (AI-guided novices)	Sens: 95–100%, Spec: 74–100%
(40–42)	Adequate image quality (novice-acquired)	80–91% of exams
(39, 40)	Reduced need for formal duplex ultrasound	Reduced by 29–53%
(43)	Carotid compressibility (ROSC detection)	Accuracy: 0.96, F1-score: 0.97
(68)	Carotid compressibility (Blood Pressure)	AUC: 0.90 (SBP<60), 0.91 (MBP<40); RMSE: 8.3 mmHg
(44)	Domain shift (IVC collapse)	Kappa fell to -0.09–0.39 on new devices
Musculoskeletal		
(45)	Wrist fracture detection	Accuracy: 0.889
(45)	Elbow fracture detection	Accuracy: 0.750
(46)	Pediatric wrist fracture detection	Sens: 1.00, Spec: ~0.90; inference <2s/sweep
(47)	Pediatric bone visualization (image enhancement)	CR improved 40%; Radiologist rating 3.9/5 vs 2.7/5
Novel Applications		

Citations	Primary Task	Key Performance Metrics
(49)	Cricothyrotomy landmark localization (Airway)	AUC >0.88; 14 FPS on iPad (real-time edge AI)
(48)	Mechanical ventilation weaning prediction (Diaphragm)	Multimodal AUC: 0.894 (vs clinical-only 0.746)

AI, artificial intelligence; POCUS, point-of-care ultrasound; ROC, receiver operating characteristic; AUC, area under the receiver operating characteristic curve; AUROC, area under the receiver operating characteristic curve; Sens, sensitivity; Spec, specificity; Acc, accuracy; NPV, negative predictive value; MD, medical doctor; COVID-19, coronavirus disease 2019; F1-Score, F1 score; LVEF, left ventricular ejection fraction; IoU, intersection over union; MAE, mean absolute error; Corr, correlation; HCM, hypertrophic cardiomyopathy; ATTR-CM, transthyretin amyloid cardiomyopathy; IVC, inferior vena cava; DVT, deep vein thrombosis; ROSC, return of spontaneous circulation; SBP, systolic blood pressure; MBP, mean blood pressure; RMSE, root mean square error; κ , Cohen's kappa; CR, contrast ratio; FPS, frames per second; TTE, transthoracic echocardiography; HF, heart failure; FAST, focused assessment with sonography for trauma. E:Expert

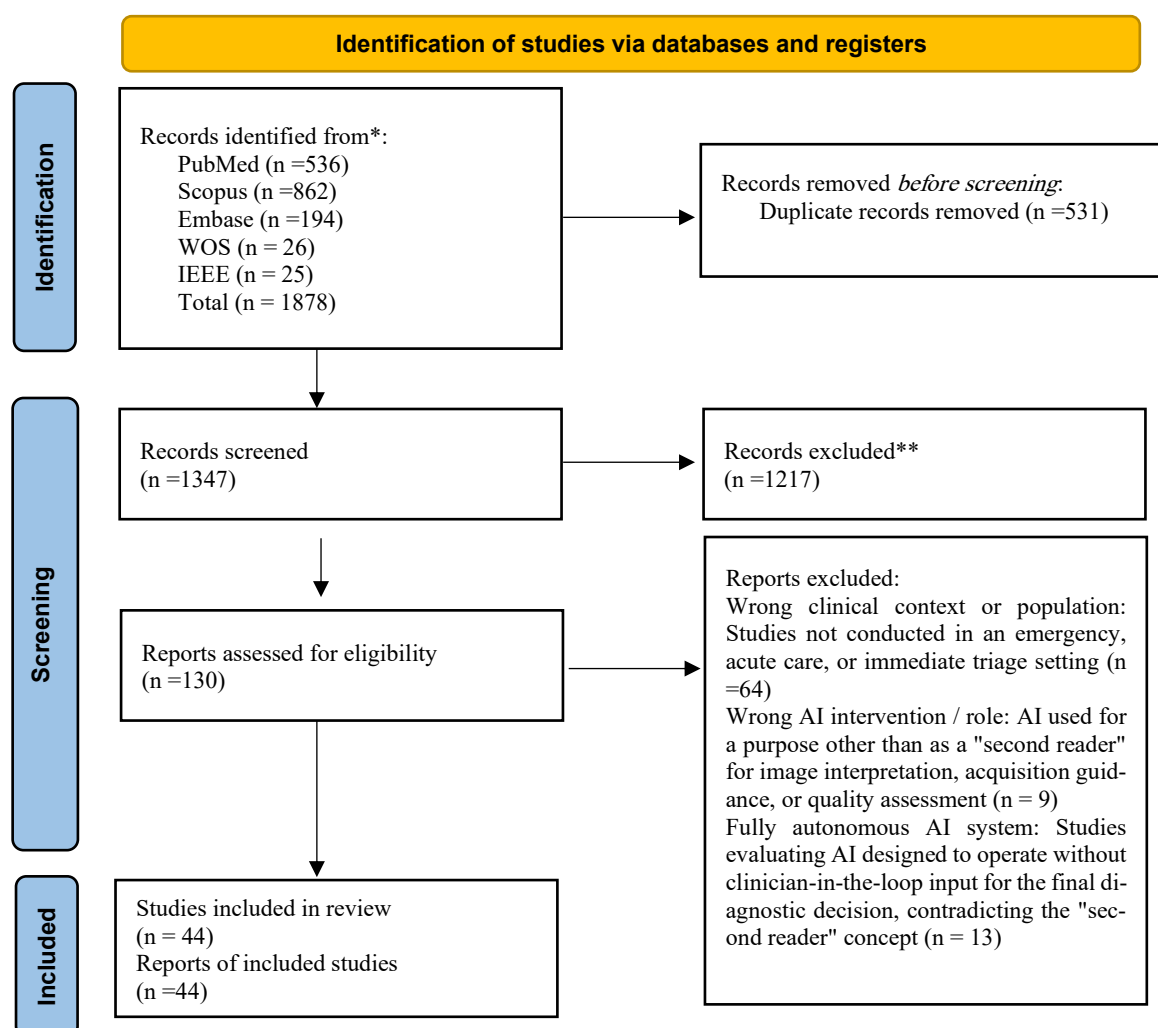


Figure 1: The PRISMA flow diagram detailing the study selection process. The diagram illustrates the number of records identified, screened, assessed for eligibility, and included in the scoping review, with reasons for exclusion at each stage.

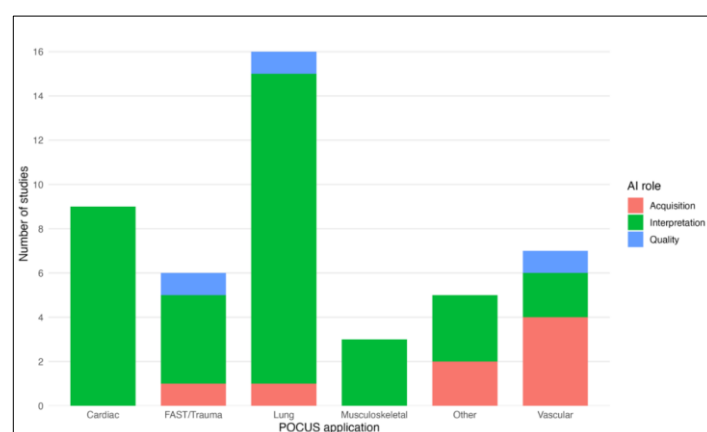


Figure 2: Distribution of POCUS applications and AI roles across the included studies. The chart summarizes the frequency of studied POCUS domains (e.g., Lung, Cardiac, Vascular) and the primary functional roles of AI (Interpretation Assistant, Acquisition Guide, Quality Assessor).

Overview of reported AI model performance across POCUS studies

Study level AUC/AUROC, sensitivity, and specificity values

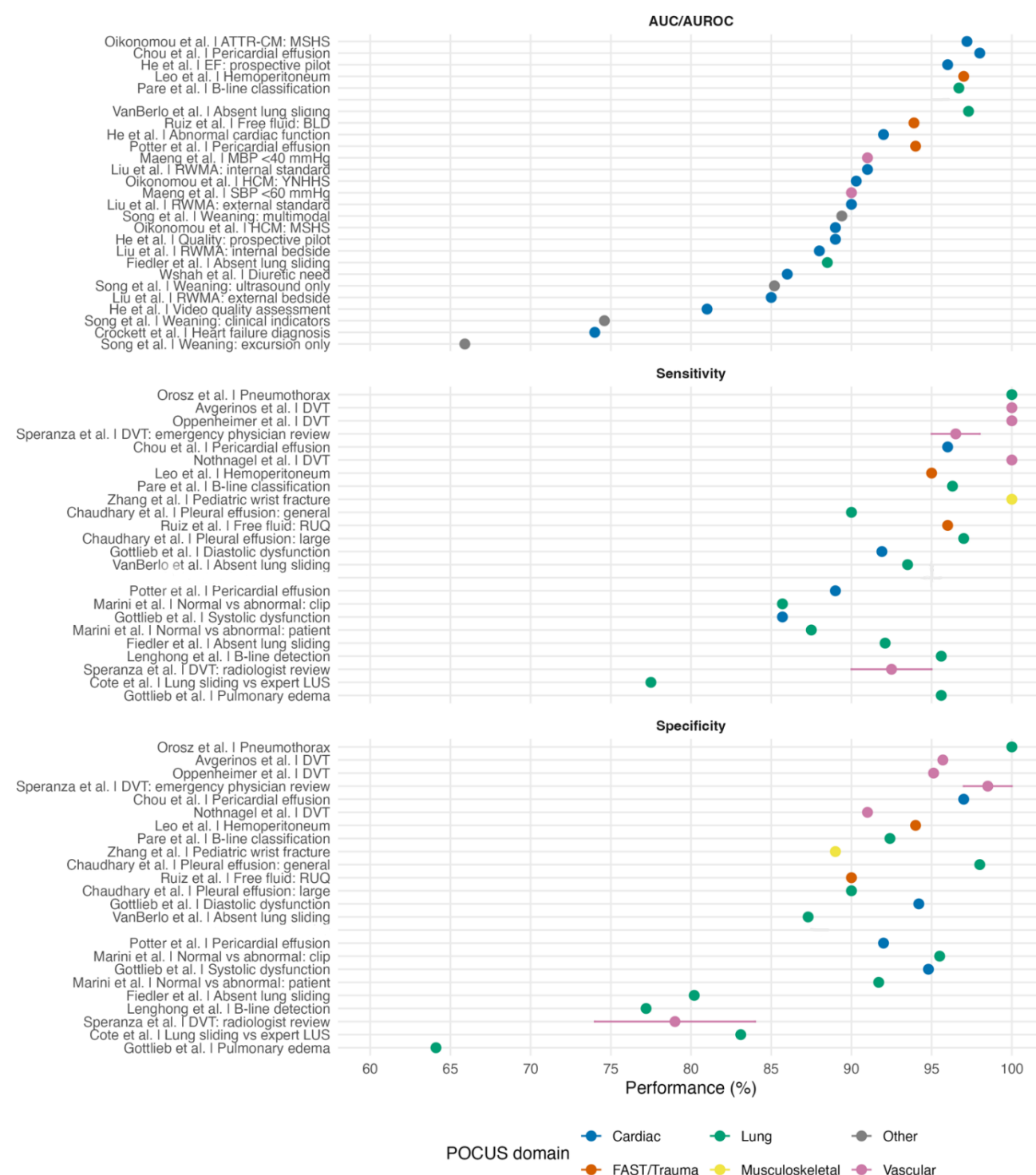


Figure 3: Overview of AI model performance metrics across studies. The visual synthesis presents key diagnostic accuracy measures (e.g., AUC, Sensitivity, Specificity) reported for various POCUS tasks, highlighting the range of reported outcomes.

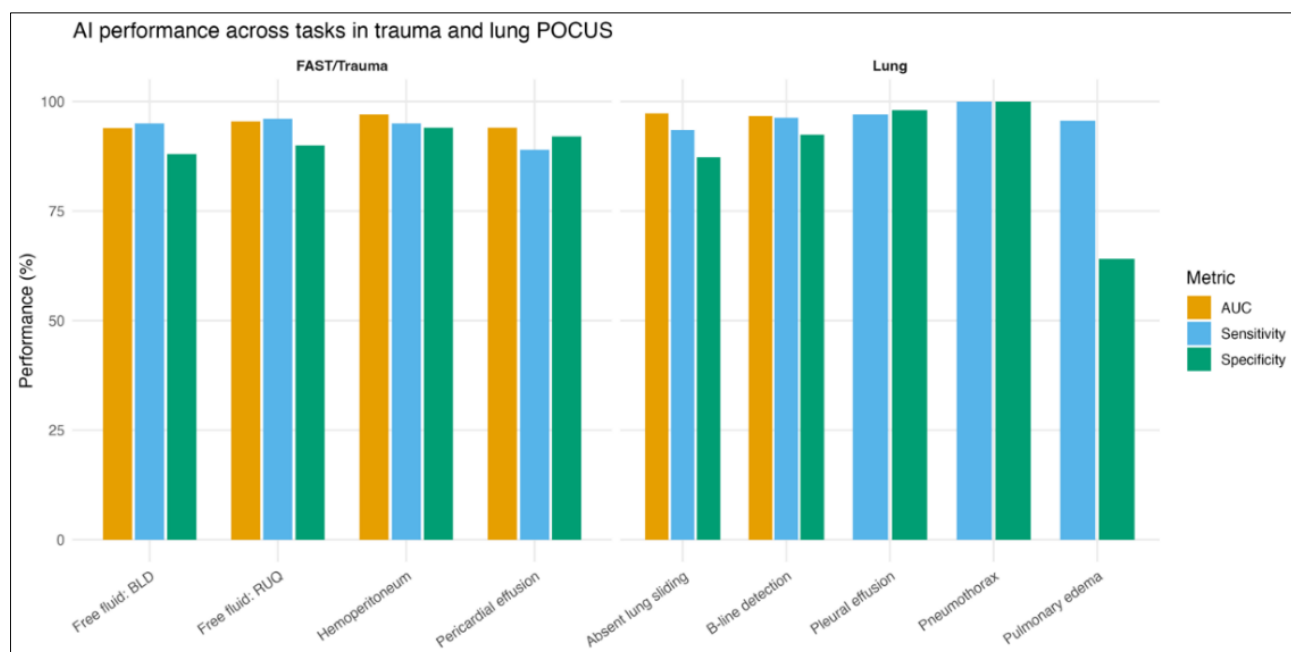


Figure 4: Comparative performance of AI models on trauma and lung ultrasound tasks. The figure juxtaposes key results for applications such as hemoperitoneum detection (FAST) and B-line or pleural effusion detection (Lung POCUS), illustrating high performance in controlled settings.

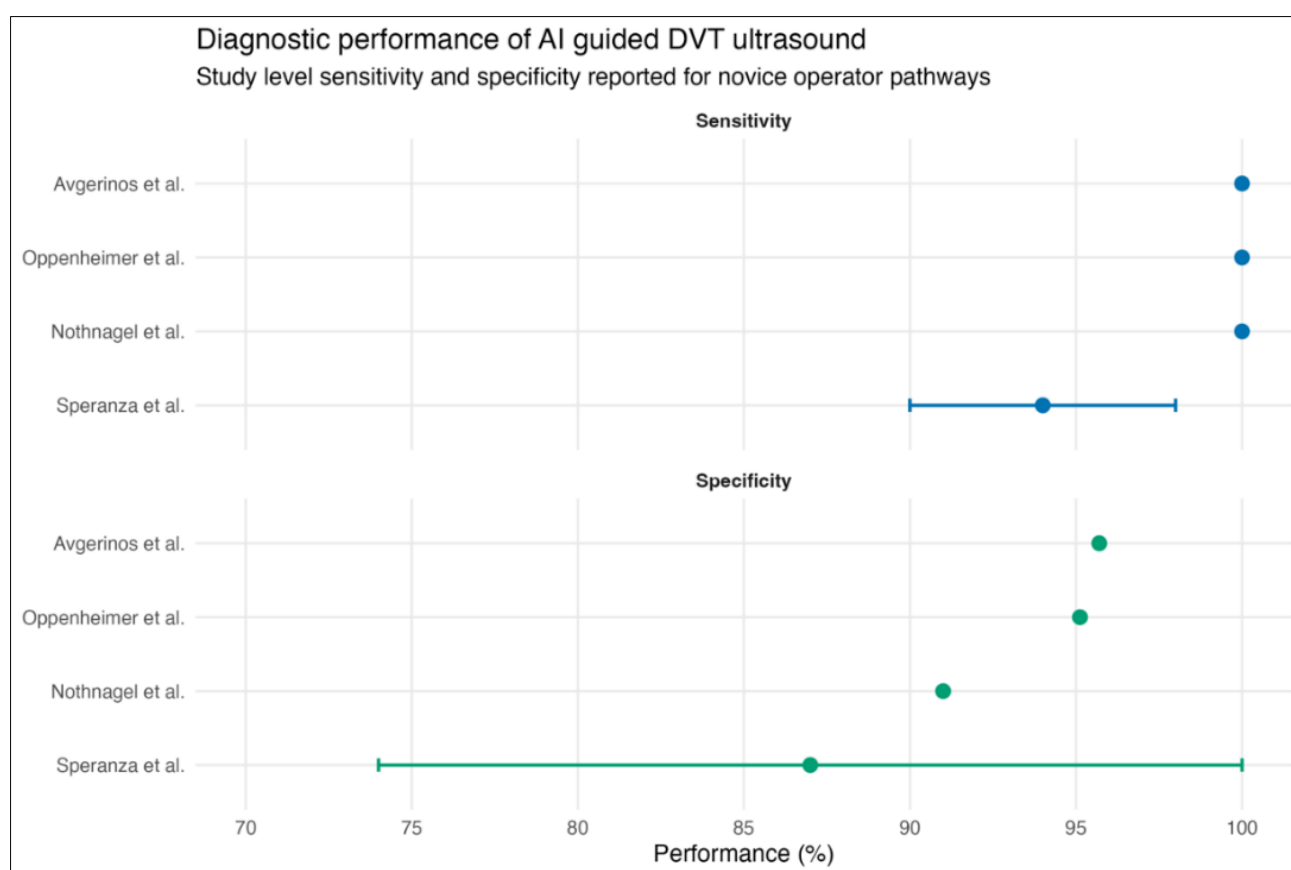


Figure 5: Reported diagnostic performance of AI-guided ultrasound for deep vein thrombosis assessment. Study-level sensitivity and specificity are shown for AI guided DVT ultrasound pathways. Horizontal lines indicate reported performance ranges where results vary across reviewers or assessment conditions.

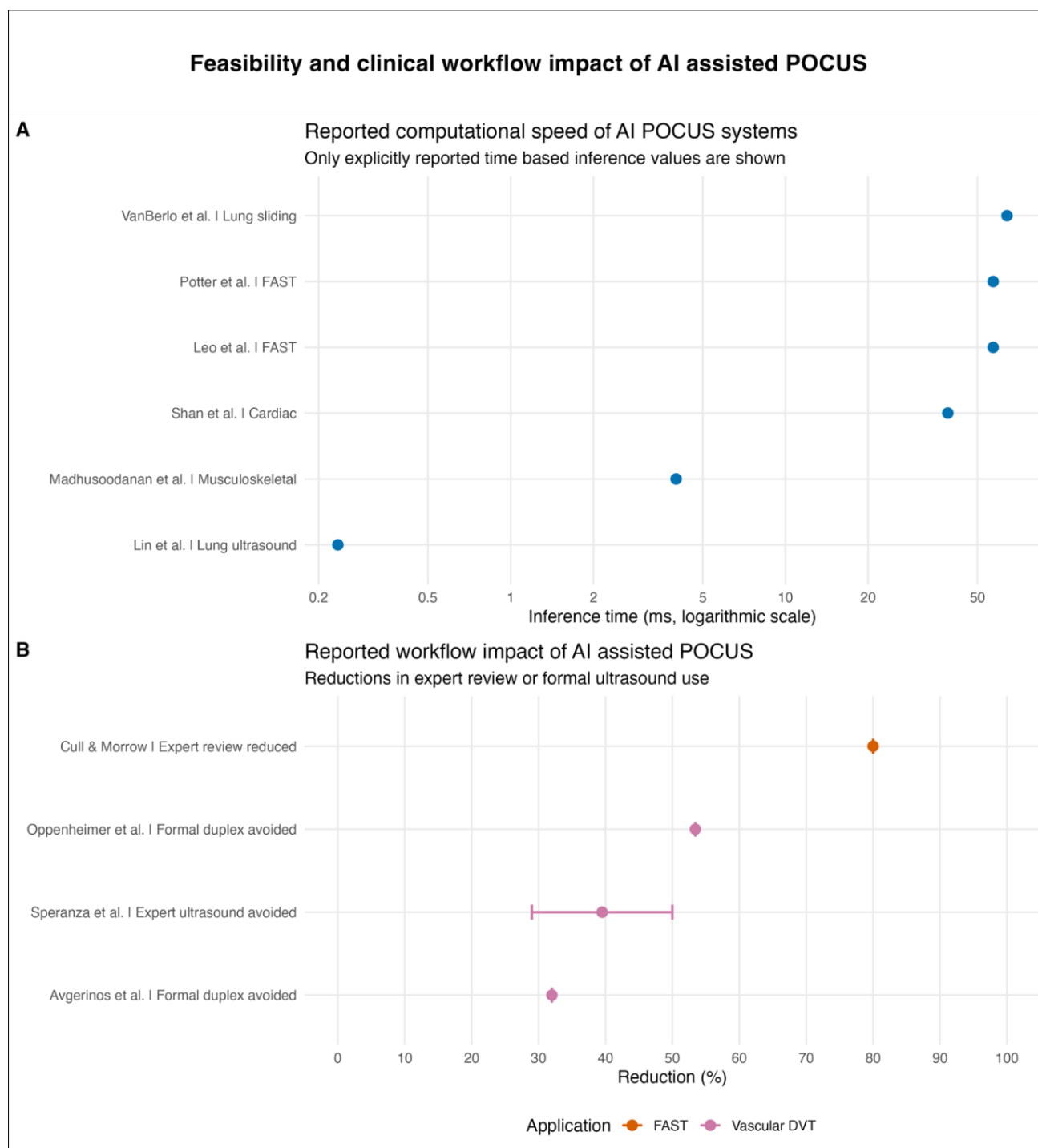


Figure 6: Computational feasibility and reported workflow impact of AI-assisted POCUS. (A) Explicitly reported inference times for AI POCUS systems, displayed on a logarithmic scale. (B) Reported reductions in expert review or formal ultrasound use following AI-assisted triage or acquisition pathways. Horizontal lines indicate reported ranges where applicable.

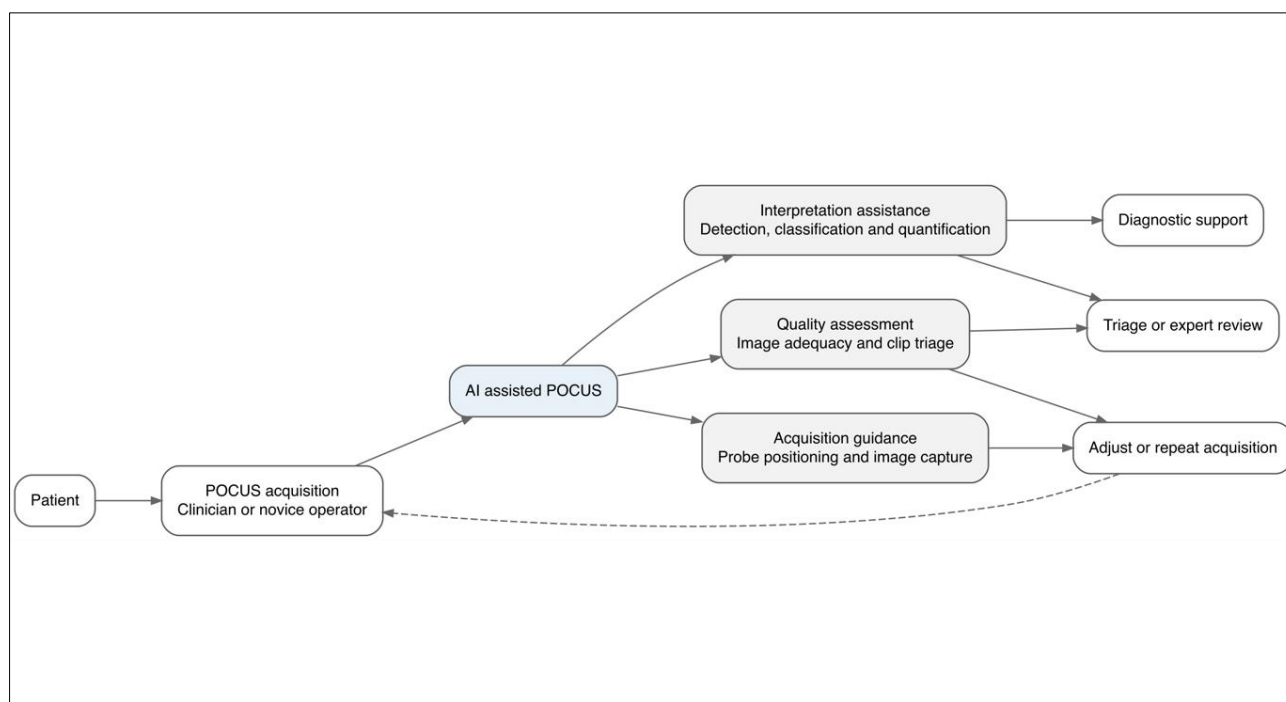


Figure 7: Conceptual diagram of an AI-augmented POCUS workflow in emergency department triage. The diagram illustrates the potential integration points for AI as a second reader, from image acquisition and quality assessment to real-time interpretation support and clinical decision-making.

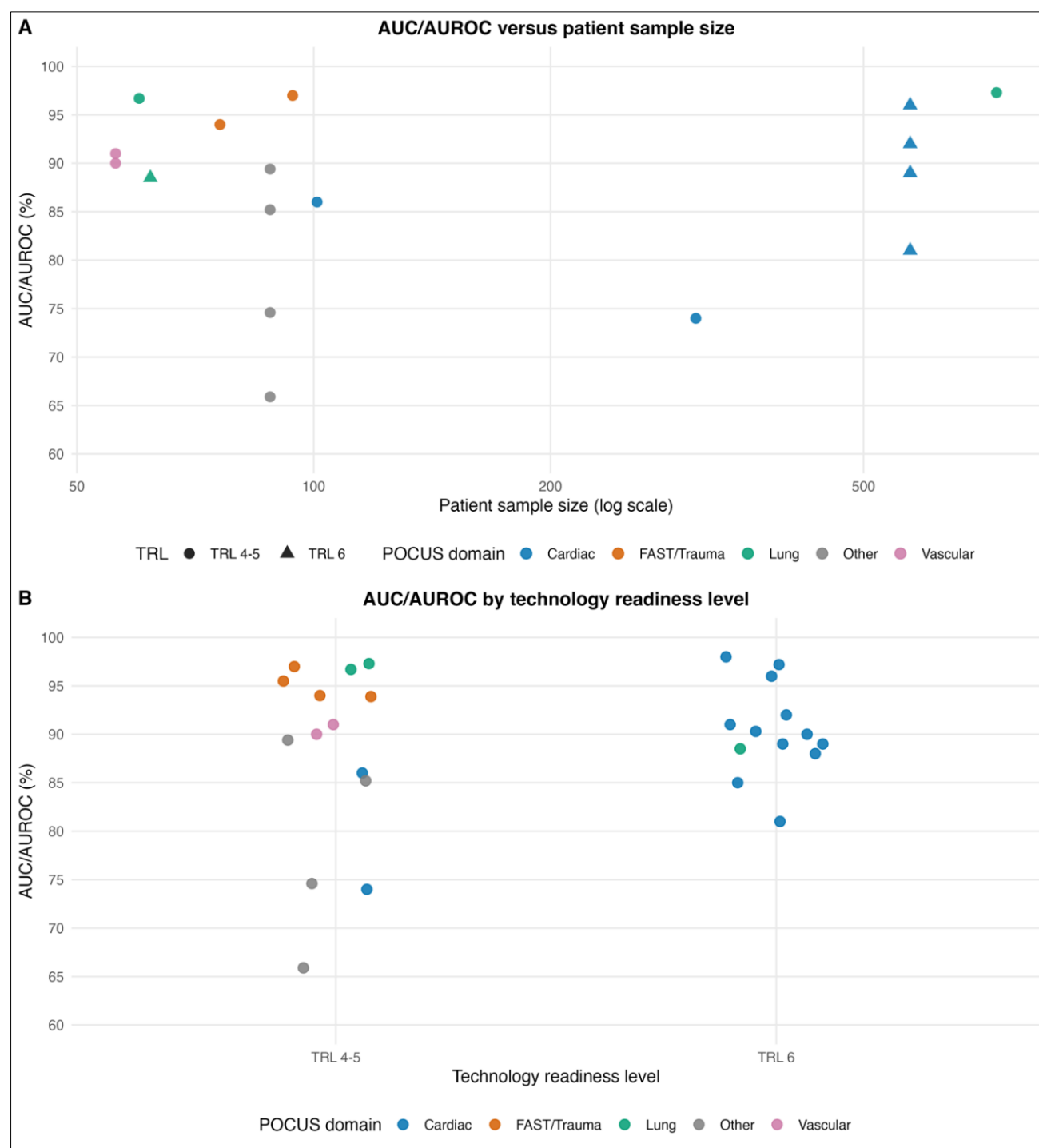


Figure 8: Relationship between reported AUC/AUROC and study characteristics. (A) Reported AUC/AUROC values according to patient sample size among studies with an identifiable patient denominator; the sample size axis is shown on a logarithmic scale. (B) Distribution of reported AUC/AUROC values according to technology readiness level. Each point represents a reported task, model, or evaluation setting; therefore, individual studies may contribute more than one observation.

Appendix

Supplementary Table 1: Search strategy in different databases

PubMed:
("Artificial Intelligence"[Mesh] OR "Deep Learning"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "Machine Learning"[Mesh] OR "Supervised Machine Learning"[Mesh] OR "convolutional neural network*" [tiab] OR "CNN"[tiab] OR "deep learning"[tiab] OR "neural network*" [tiab] OR "computer vision"[tiab] OR "AI"[tiab])
AND
("Ultrasonography"[Mesh] OR "Point-of-Care Systems"[Mesh] OR "Echocardiography"[Mesh] OR "Ultrasonography, Doppler"[Mesh] OR "POCUS"[tiab] OR "point-of-care ultrasound"[tiab] OR "point of care ultrasound"[tiab] OR "bedside ultrasound"[tiab] OR "focused ultrasound"[tiab] OR "E-FAST"[tiab] OR "eFAST"[tiab] OR "FAST exam"[tiab] OR "Focused Assessment with Sonography for Trauma"[tiab] OR "bedside echocardiography"[tiab] OR "focused echocardiography"[tiab] OR "lung ultrasound"[tiab] OR "pleural ultrasound"[tiab] OR "vascular ultrasound"[tiab])
AND
("Emergency Service, Hospital"[Mesh] OR "Triage"[Mesh] OR "Emergency Medicine"[Mesh] OR "Point-of-Care Systems"[Mesh] OR "emergency department"[tiab] OR "emergency departments"[tiab] OR "ED"[tiab] OR "EDs"[tiab] OR "triage"[tiab] OR "triaged"[tiab] OR "emergency room"[tiab] OR "ER"[tiab] OR "accident and emergency"[tiab] OR "resuscitation"[tiab] OR "critical care"[tiab])
Scopus:
(TITLE-ABS-KEY ("convolutional neural network*" OR cnn OR "deep learning" OR "neural network*" OR "computer vision" OR ai OR "machine learning" OR "supervised learning" OR "image classification" OR "object detection"))
AND
(TITLE-ABS-KEY ("point-of-care ultrasound" OR pocus OR "bedside ultrasound" OR "focused ultrasound" OR "E-FAST" OR efast OR "FAST exam" OR "Focused Assessment with Sonography for Trauma" OR "bedside echocardiography" OR "focused echocardiography" OR "lung ultrasound" OR "pleural ultrasound" OR "vascular ultrasound" OR ultrasonography OR echocardiography))
AND
(TITLE-ABS-KEY ("emergency department" OR "emergency medicine" OR "emergency room" OR er OR ed OR triage OR triaged OR "accident and emergency" OR resuscitation OR "critical care" OR "point-of-care"))
EMBASE:
(('artificial intelligence'/exp OR 'machine learning'/exp OR 'deep learning'/exp OR 'convolutional neural network':ti,ab OR 'cnn':ti,ab OR 'computer vision':ti,ab OR 'ai':ti,ab))
AND
('point of care testing'/exp OR 'pocus':ti,ab OR 'point-of-care ultrasound':ti,ab OR 'point of care ultrasound':ti,ab OR 'bedside ultrasound':ti,ab OR 'focused ultrasound':ti,ab OR 'e-fast':ti,ab OR 'fast exam':ti,ab)
AND
('emergency health service'/exp OR 'triage'/exp OR 'emergency department':ti,ab OR 'ed':ti,ab OR 'triage':ti,ab OR 'emergency room':ti,ab OR 'er':ti,ab)
IEEE:
("convolutional neural network" OR "deep learning" OR "computer vision" OR CNN OR AI OR "machine learning" OR "neural network" OR "artificial intelligence") "All Metadata"
AND
("point-of-care ultrasound" OR POCUS OR "bedside ultrasound" OR "E-FAST" OR "FAST exam" OR "focused ultrasound" OR "lung ultrasound" OR "vascular ultrasound" OR "echocardiography" OR "ultrasonography") "All Metadata"
AND
("emergency department" OR ED OR triage OR "emergency room" OR ER OR "emergency medicine" OR resuscitation OR "critical care") "All Metadata"
WOS
TS=(("artificial intelligence" OR "AI" OR "deep learning" OR "neural network*" OR "convolutional neural network*" OR "CNN" OR "machine learning" OR "supervised machine learning" OR "computer vision"))
AND
TS=(("point-of-care" NEAR/2 (ultrasound OR ultrasonography OR echography)) OR "POCUS" OR ("bedside" NEAR/2 (ultrasound OR echocardiography)) OR ("focused" NEAR/2 (ultrasound OR echocardiography)) OR "E-FAST" OR "eFAST" OR "FAST exam" OR "Focused Assessment with Sonography for Trauma" OR "lung ultrasound" OR "pleural ultrasound" OR "vascular ultrasound" OR echocardiography OR "Ultrasonography, Doppler"))
AND
TS=(("emergency department*" OR "ED" OR "EDs" OR "emergency room" OR "ER" OR "accident and emergency" OR triage OR triaged OR resuscitation OR "critical care" OR "emergency medicine"))

Supplementary Table 2: Inclusion and exclusion criteria for study selection

Category	Inclusion Criteria	Exclusion Criteria
Study Focus	Focuses on the development, validation, or implementation of AI/ML for assisting in the interpretation, acquisition, or quality assessment of POCUS for clinical decisions in an ED or similar urgent care setting.	Studies where AI is used for non-emergency, comprehensive ultrasounds (e.g., routine radiology-performed echocardiography) or for guiding procedures (e.g., vascular access).
Population	Studies conducted on human patients in an emergency or urgent care setting, or using datasets derived from these settings.	Studies conducted solely on healthy volunteers, in non-acute outpatient clinics, or using only simulated or phantom data without real patient validation.
AI Intervention	AI model must analyze the raw or processed POCUS image or video clip. Outcomes must include diagnostic accuracy, image quality assessment, impact on clinical decision-making/triage, or workflow efficiency. Includes studies where AI enables remote expert review by ensuring adequate image acquisition.	Studies where AI is only used for image compression, archival, enhancement without diagnostic intent, or non-clinical tasks (e.g., billing, reporting).
"Second Reader" Role	The AI must function to support a human clinician's decision-making. This includes systems where the AI provides an input that is part of a clinical workflow (e.g., flags a finding, guides a novice, assesses quality for remote read).	Studies where the AI is presented as a fully autonomous diagnostic system with no intended human-in-the-loop for the final clinical decision at triage.
Publication Type and Date	The AI must function to support a human clinician's decision-making. This includes systems where the AI provides an input that is part of a clinical workflow (e.g., flags a finding, guides a novice, assesses quality for remote read).	Case reports, editorials, letters, conference papers, narrative reviews, protocols, and studies not published in English. Studies published before 2020, unless they are seminal foundational works identified during the search.