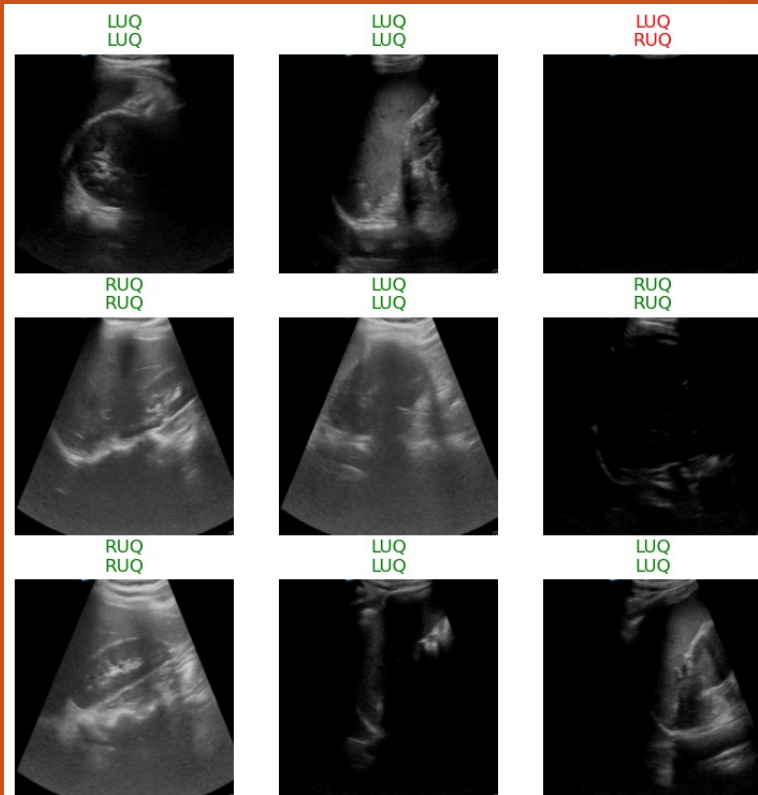


A deep learning abdominal trauma ultrasound model demonstrated comparable performance when validated on retrospective and prospective datasets.

Prediction:
Target:



Prospective Validation of a Deep Learning View Classifier for Pediatric Focused Assessment with Sonography for Trauma (FAST)

Nathan Velarde¹, Newton Addo², Katie Gutierrez¹, Kevin Li², Margaret Lin-Martore, MD², Aaron Kornblith, MD²

¹School of Medicine, University of California San Francisco

²Department of Emergency Medicine, University of California San Francisco



University of California
San Francisco

Background

- Lack of FAST expertise among clinicians for injured children limits its use in identifying life saving information¹⁻³
- Deep learning anatomic FAST view classifier developed and validated on retrospective studies
- The overall accuracy of 99.1%, decreased to 88.5% when discriminating right upper quadrant (RUQ) and left upper quadrant (LUQ) views⁴
- Retrospective image labeling may introduce bias - e.g. expert labeling of RUQ vs LUQ abdominal views
- Prospective validation is needed to determine its predictive performance.

Objective: The aim of this study was to validate the sensitivity of a deep learning FAST view classifier in discriminating between the right upper quadrant and left upper quadrant abdominal views by testing it on a prospective validation set and comparing its performance to the original retrospective validation set.

Methods

- Prospective convenience sample
- Inclusion: volunteer students undergoing routine FAST education at UCSF from September 2022 – February 2023
- Primary outcome: FAST view classification sensitivity by majority clip
- Ground truth determined by real-time image labeling at time of acquisition (compared to expert label)

Results

- 129 total upper abdominal clips from 21 participants.
- The sensitivity for detecting LUQ clips was 88.7 % (95% CI: 76.3 - 94.2%) on the prospective validation set and 86.4% (95% CI: 66.6 - 95.3%) on the retrospective validation set.
- The sensitivity for detecting RUQ clips was 81.0% (95% CI: 69.1 - 89.1%) for the prospective validation set and 68.0% (95% CI: 48.4 - 82.7%) for the retrospective validation set.

Prospective Validation Dataset

Target	Prediction			
		LUQ	RUQ	Any Other*
LUQ		63	5	3
RUQ		10	47	1

Retrospective Validation Dataset

Target	Prediction			
		LUQ	RUQ	Any Other*
LUQ		19	3	0
RUQ		6	17	2

*Any other refers to cardiac, suprapubic, and lung views

Discussion

- A deep learning FAST model trained on retrospective data maintains comparable performance on prospective and retrospective validation.
- This suggests that the costs of using prospective validation instead of retrospective validation for the initial validation of deep learning-based FAST technologies may outweigh the potential benefits.
- Additional improvements to the model's performance may require alternative approaches like removing low quality studies from the test set or more detailed annotation tasks

Acknowledgements: The research reported here was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number K23HD110716-01.