



Evaluating the Clinical Utility and Diagnostic Impact of AI Triage for Incidental Pulmonary Embolism (IPE)

Aawez Mansuri, MS, Emory University; Mohammedreza Chavoshi, MD; Theo Dapamede, MD, PhD; Dan Cohen, MD; Ninad Salastekar, MD; Wasif Bala, MD; John Moon, MD; Hanssen Li, MD; Chad Robichaux, MPH; Bardia Khosravi, MD, MPH, MHPE; Frank Li, PhD; Rohan Issac, MS; Beatrice Brown-Mulry; Janice Newsome, MD; Judy Gichoya, MD, MS, FSIIM; Hari Trivedi, MD

Introduction/Background

Incidental pulmonary embolism (IPE) is often missed on routine contrast-enhanced CTs done for reasons other than PE detection. We assessed an FDA-cleared AI algorithm at our institution by (1) measuring its diagnostic accuracy against report-based ground truth, and (2) evaluating its clinical impact based on follow-up PE-protocol exams in patients with AI-flagged and unflagged IPE cases.

Methods/Intervention

We retrospectively analyzed 37,272 contrast-enhanced chest and abdominal CT reports over one year, using GPT-4o to automatically label pulmonary embolism (PE) status. GPT-4o was previously validated, achieving 100% sensitivity and 96.1% specificity on 500 radiologist-annotated reports. To avoid including known PE cases, we excluded exams with a PE-protocol CT performed ≤ 7 days before or ≤ 24 hours after the index study. To assess clinical impact, we tracked patients who underwent a PE-protocol scan within 30 days after an IPE exam, and compared the AI model's results with radiology report labels to evaluate diagnostic agreement and potential influence on patient care.

Results/Outcome

Of the 37,272 studies evaluated, 538 cases were diagnosed with IPE. The IPE model detected 354 of these cases, yielding a sensitivity of 65.8%, specificity of 99.8%, and an F1 score of 0.72. Among the 786 patients who underwent a subsequent PE-protocol exam, 40 had their IPE flagged by the AI model, while 12 cases went undetected. Of the 106 patients confirmed positive for PE on follow-up imaging, 28 were true positives (TP) detected by the AI. Among the remaining cases, 651 patients were negative for PE, with 639 true negatives (TN) identified, and 12 false positives (FP) flagged by the AI. Thus, while the tool captured the majority of IPE cases, missed detections still represent a clinically meaningful concern.

Conclusion

The AI triage system shows excellent specificity but only moderate sensitivity, leaving a consequential subset of IPEs undetected—some with confirmed emboli on subsequent reports. These findings highlight the need for further algorithm improvement and continued clinical oversight.

Statement of Impact

AI triage tools for IPE detection can aid in identifying otherwise overlooked pulmonary emboli and potentially expedite clinical intervention. However, their limitations reinforce the importance of radiologist review and further model improvement to ensure reliable clinical integration.

Table 1: AI IPE Detection and Corresponding PE Follow-Up Results

AI IPE Model Result	IPE Reported by Radiologist	Positive	Negative	Possible	Case Count
Negative	Yes	6	5	1	12
	No	70	628	27	725
	Possible	2	6	1	9
Positive	Yes	18	8	0	26
	No	2	3	0	5
	Possible	8	1	0	9
Total	—	106	651	29	786

Table 1: AI IPE Detection and Corresponding PE Follow-Up Results

Keywords

Incidental Pulmonary Embolism; Large Language Models; Real World Validation