GLOBAL AI CONFERENCE LONDON · 2026

IRB CERIM 2512-520

No funding · No conflicts of interest

Balancing alert burden and diagnostic recovery: A Multicenter Evaluation of Five Concurrently Deployed AI Triage Tools in Emergency Radiology

Mohamed Hedi Bouchhioua[1] Antoine Noguero[2] Guillaume Herpe[2]

[1] Université Paris Cité, UFR de Médecine · [2] CHU de Poitiers, Radiologie d'Urgence & Intelligence Artificielle

4,421 Imaging examinations *included (4,004 patients)*

38 Missed diagnoses recovered by AI

0.53 False-positive alerts/day *95% CI: 0.40–0.71*

7.9 Recovered misses/10 false alerts

BACKGROUND & OBJECTIVE

- **High-volume emergency radiology** carries a persistent **risk of missed critical diagnoses**, particularly for time-sensitive conditions
- **AI triage tools prioritize urgent cases**, yet most evidence focuses on isolated tools or controlled settings. The combined impact of multiple concurrent AI modules in real workflows is poorly characterized.

We evaluated five simultaneously deployed AI triage tools across three emergency radiology centers, quantifying recovered missed diagnoses against false-positive alert burden — and propose a novel operational efficiency metric.

METHODS

1. **Design** : Retrospective multicenter study — 90-day period (March–May 2025)
2. **Population** : 4,004 patients · median age 58 years (IQR 41–72) · 51.2% male · 476/4,897 exams excluded
3. **Centers :** 3 French emergency radiology departments 1 tertiary and 2 secondary · 25 senior radiologists
4. **Platform** : 5 CE-marked/FDA-cleared algorithms deployed via Incepto (Paris) · *AI outputs were retrospective*
5. **Adjudication :** Discordant cases - 2 blinded senior radiologists (*≥10 yrs*); consensus for disagreements
6. **Efficiency metric :** recovered misses/(FP alerts × 10) *calculated per module and overall*

AI - Algorithms Deployed

- qEr v1.2 — ICH (Qure, Mumbai)
- SmartChest v1.0 — Pneumothorax (Milvue, Paris)
- cinaIPE v1.1 — Incidental PE (Avicenna)
- SmartUrgence — Fracture (Milvue, Paris)
- CinaLVO v2.0 — LVO (Avicenna)

STUDY CENTERS

Tertiary	*1,601 patients · 1,983 exams*
Secondary 1	*944 patients · 1,343 exams*
Secondary 2	*1,459 patients · 1,571 exams*

Five AI Modules — Per-Module Results

Intracranial hemorrhage (ICH)
Non-contrast head CT

682 Exams · 115 Alerts · 28 FP

2 recovered *radiologist misses*

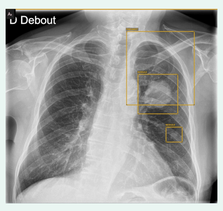
Sens 88.8% · Spec 95.2% · PPV 75.7% · NPV 98.1%

Pneumothorax
Chest radiograph

595 Exams · 45 Alerts · 5 FP

10 recovered *radiologist misses*

Sens 88.9% · Spec 99.1% · PPV 88.9% · NPV 99.1%



Pulmonary embolism (IPE)
Contrast-enhanced CT

414 Exams · 13 Alerts · 8 FP

2 recovered *radiologist misses*

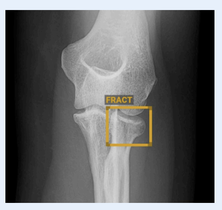
Sens 100% · Spec 98.0% · PPV 38.5% · NPV 100%

Fracture
Extremity radiograph

2430 Exams · 935 Alerts · 5 FP

21 recovered *radiologist misses*

Sens 97.9% · Spec 98.7% · PPV 95.0% · NPV 98.9%

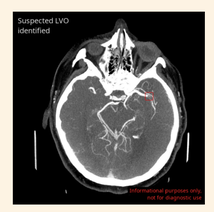


Large vessel occlusion (LVO)
Head/neck CTA

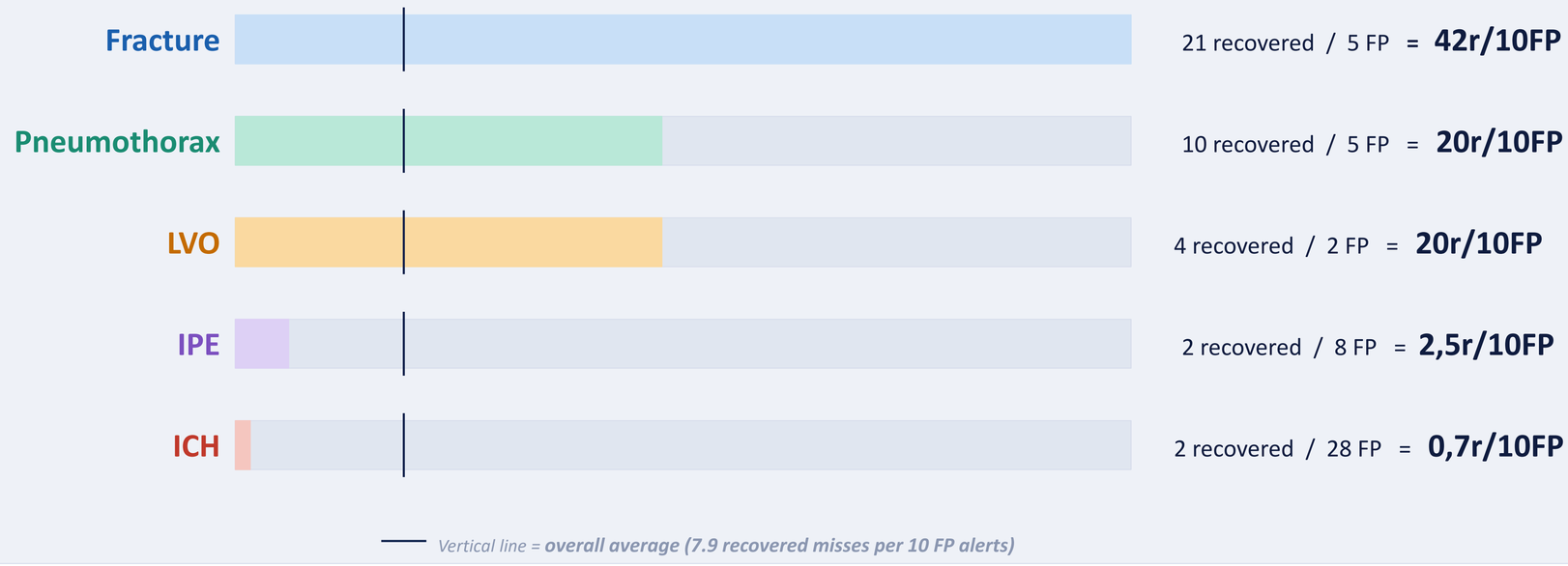
300 Exams · 16 Alerts · 2 FP

4 recovered *radiologist misses*

Sens 93.3% · Spec 99.3% · PPV 87.5% · NPV 99.6%



OPERATIONAL EFFICIENCY — RECOVERED MISSED DIAGNOSES (r) PER 10 FALSE-POSITIVE ALERTS (FP)



Module-Level Diagnostic Performance & Alert Yield

Module (Modality)	Exams	Alerts	TP	FP	FN	Sensitivity [95%CI]	Specificity [95%CI]	Recovered	Eff/10 FP
ICH (Head CT)	682	115	87	28	11	**88.8%** [80.8–94.3]	**95.2%** [93.5–96.9]	2	**0.7**
Pneumothorax (CXR)	595	45	40	5	5	**88.9%** [76.0–96.3]	**99.1%** [98.3–99.8]	10	**20.0**
IPE (Contrast CT)	414	13	5	8	0	**100%** [47.8–100]	**98.0%** [96.8–99.4]	2	**2.5**
Fracture (XR)	2,430	935	915	5	21	**97.9%** [95.6–98.9]	**98.7%** [96.9–100]	21	**42.0**
LVO (CTA)	300	16	14	2	1	**93.3%** [80.7–100]	**99.3%** [98.3–100]	4	**20.0**
Total / Overall	**4,421**	**1,124**	**1,061**	**48**	**38**	**96.6% [95.5–97.7]**	**98.8% [98.4–99.1]**	**38**	**7.9**

38

Clinically relevant missed diagnoses recovered across all 5 modules over 90 days *including life-threatening conditions (ICH, PE, LVO, PTX) and urgent fractures — at a manageable false-positive burden of 0.53 FP alerts/day (95%CI: 0.40–0.71).*

7.9

The proposed "recovered misses per 10 FP alerts" metric offers a practical, module-level indicator for the benefit-to-burden balance — *informing AI governance, calibration decisions, and prospective evaluation in high-throughput emergency settings.*