

AI enabled CT head triage: A service evaluation to reduce outsource expenditure

W Hughes, S Gangoli



Background

Approximately £600,000 per year is spent on out-of-hours (OoH) non-contrast CT head (NCCTHs) report outsourcing. Internal estimates suggested 80% of scans returned non-urgent findings. Financial modelling indicated that an effective AI triage tool could generate savings up to £450,000 annually by deferring low-risk cases for in-house reporting.

Aims

This evaluation assesses the feasibility of using a CE-marked deep learning tool (Annalise.AI CT Brain) capable of detecting 130 findings to triage OoH NCCTHs and reduce unnecessary outsourcing of non-urgent studies.

Methods

Data Collection.

2,118 anonymised studies included, limited to adult (≥18) out-of-hours exams

Ground Truthing

Expert radiologists classified each of the 130 possible findings as No Findings (0), Non-Urgent/Technical (1) or Urgent (2). Report texts were manually reviewed and assigned one of three labels (0, 1, or 2)

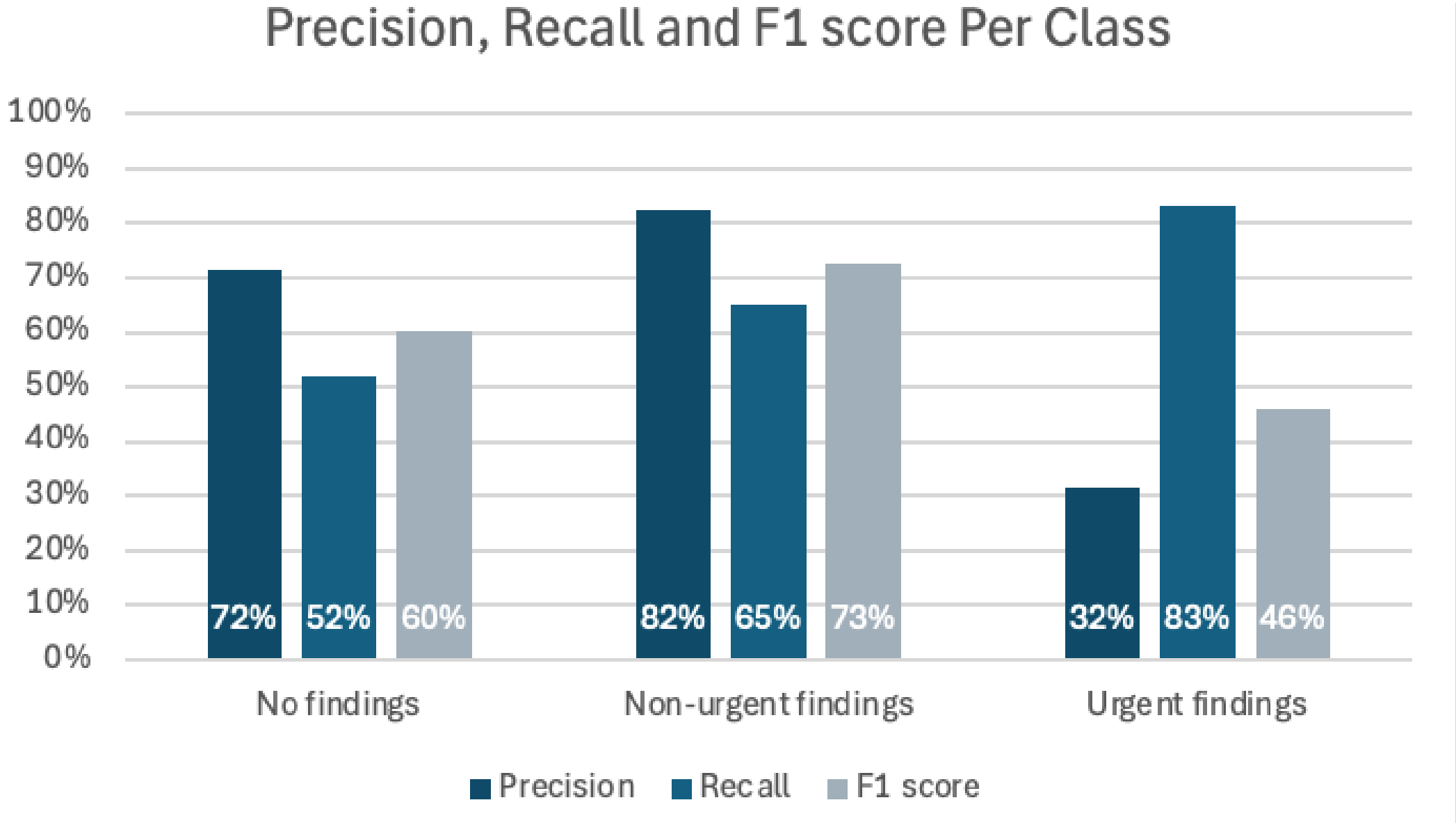
Performance Metrics

- Sensitivity, Specificity
- Positive Predictive Value (PPV)
- Negative Predictive Value (NPV)
- Per Class Precision, Recall and F1 scores for performance
- Cohen’s weighted kappa used to assess agreement across ordinal classifications.

Results

'Urgent' class Performance Metrics		
Metric	Value	95% CI
Sensitivity	83.3%	78.2 – 87.3
Specificity	34.0%	30.6 – 37.6
Positive Predictive Value (PPV)	31.6%	28.2 – 35.2
Negative Predictive Value (NPV)	90.3%	88.6 – 91.8
Accuracy	64.4%	62.3 – 66.4
Cohen's Kappa	0.464	0.437–0.491

Error classifications	
Urgent false positives	464 studies (21.9%) where AI incorrectly flagged as urgent
Routine false positives	168 studies (7.9%) where AI incorrectly flagged as non-urgent
Urgent false negatives	43 studies (2.0%) where AI misrepresented an urgent finding
Routine false negatives	80 studies (3.8%) where AI misrepresented a non-urgent finding



Insights

Key Messages

This evaluation found that the AI tool demonstrated **high sensitivity** (83.3%) and **negative predictive value** (90.3%), indicating strong technical potential as a triage aid. However, **modest specificity** (34.0%) and **positive predictive value** (31.6%) contributed to a **high false-positive rate** (21.9%), raising concerns for standalone-use.

Based on analysis of the results and subsequent stakeholder discussions, several potential barriers to real-world implementation were identified:

- **Workflow considerations:** False positives and deferred cases may increase, rather than decrease, clinical workload unless tightly integrated with existing reporting pathways.
- **Trust and interpretability:** Without structured training or explainable outputs, there is a risk that non-radiologist users may lack confidence in AI-driven decisions or rely on them inappropriately.
- **Operational readiness:** Realising projected cost savings likely depends on dedicated infrastructure for follow-up, audit, and governance – which may be lacking in some settings.

Next Steps

Deploy the tool in a prospective ‘shadow mode’ to assess real-time performance without influencing clinical decisions.

Conduct structured user feedback studies to better understand attitudes, behaviours, and confidence in interpreting AI outputs.

Define governance and escalation protocols to support safe, accountable use of AI triage in clinical workflows.

Develop and test workflow-aligned integration models that reduce rather than add to reporting burden.

Reassess cost-benefit projections in light of practical implementation challenges.