

Real-World Deployment of AI Assisted CT Brain Triage: A Multi-Site Performance Evaluation

D. Green, P. Perera, S. Gangoli

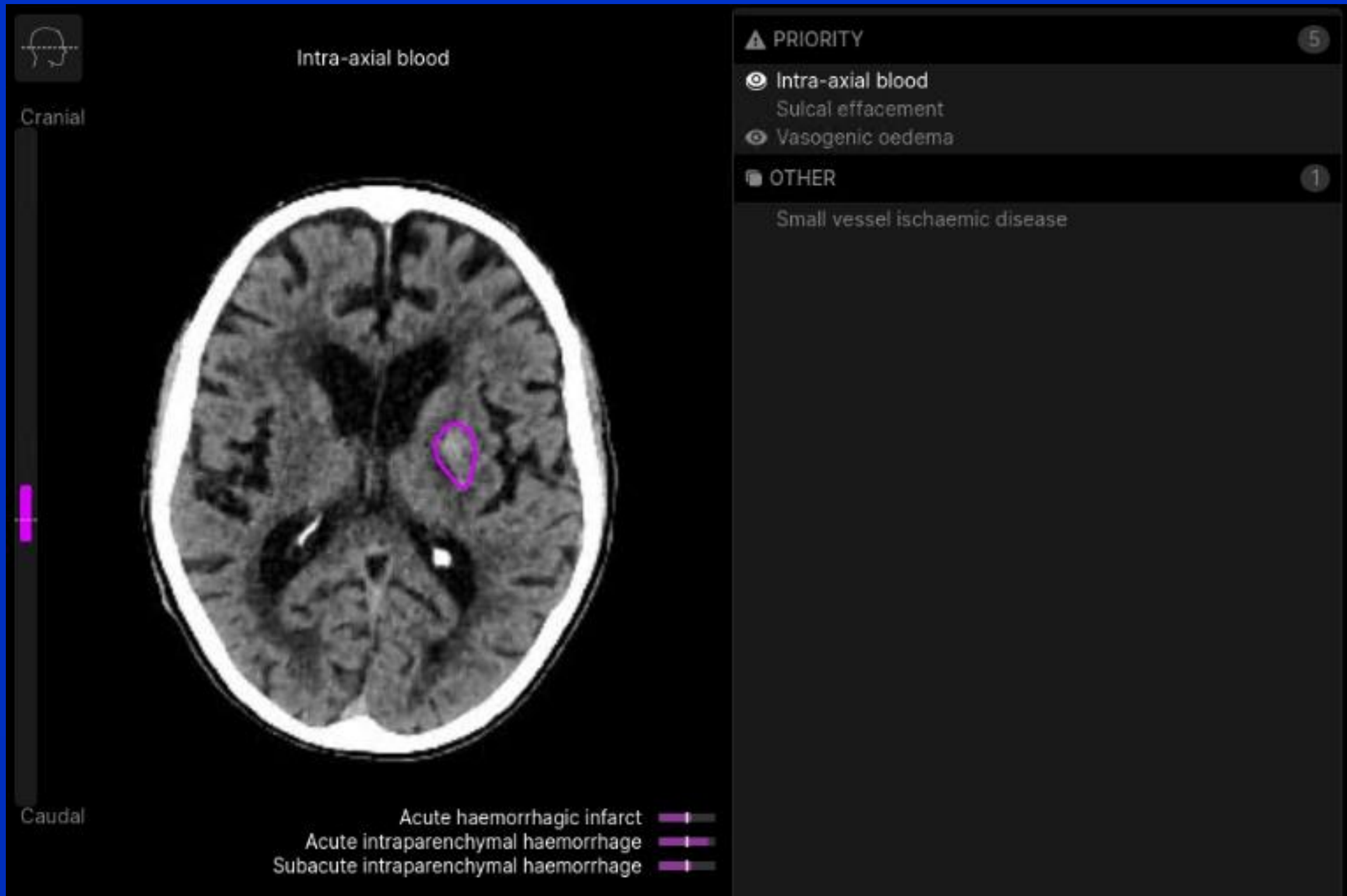


The Problem:

University Hospitals Sussex NHS Foundation Trust serves approximately 1.7 million people and performs ~9000 CT Brain studies out of hours per year at a cost of ~£600,000. There is some variability across the Eastern and Western areas of the trust, with the former containing a tertiary neurosciences centre.

Could an AI triage tool be used to identify low risk, likely normal studies and defer their reporting to more cost-effective hours of operation?

The Tool: Annalise.ai CTB Enterprise

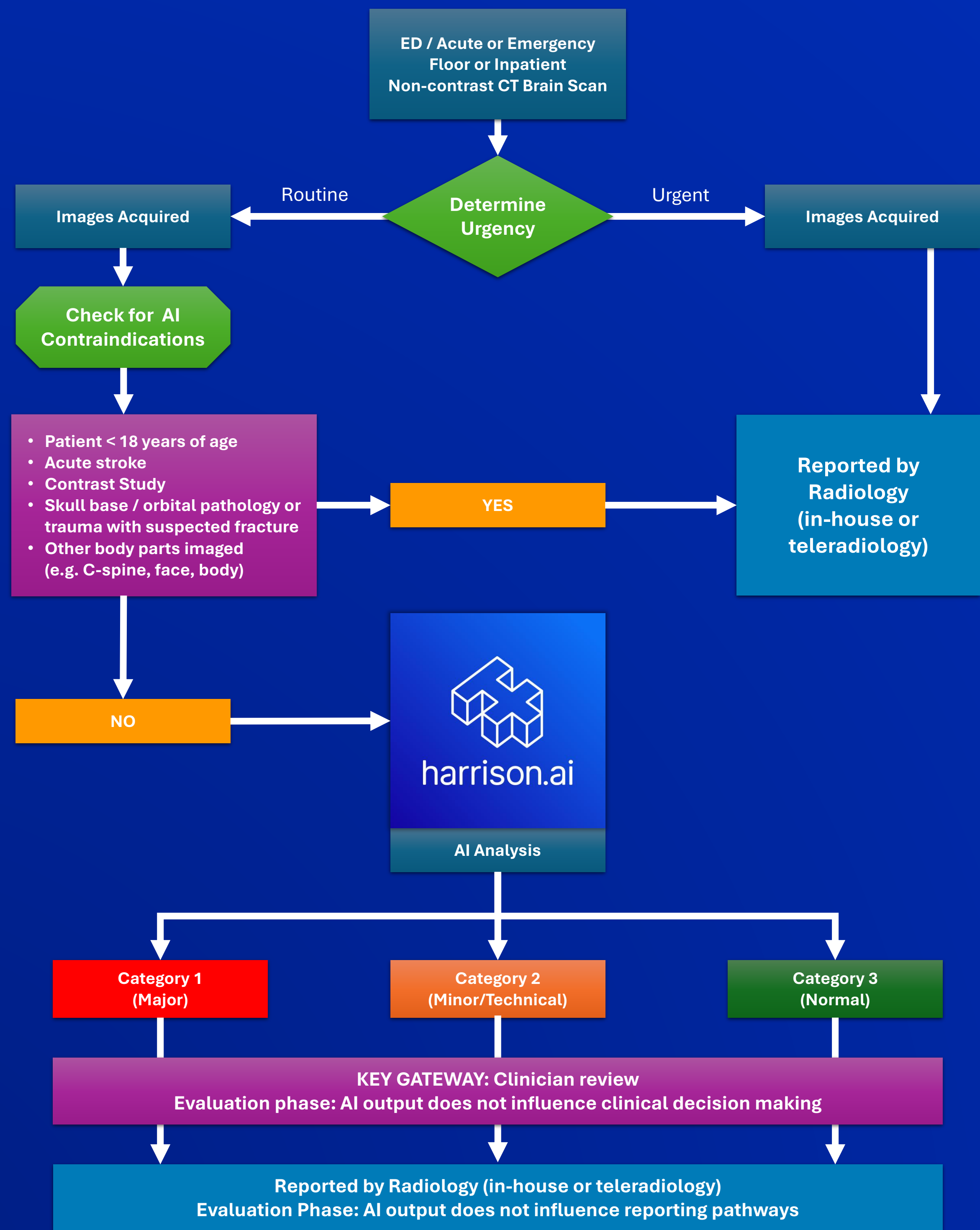


- Ground truth based on hand labelled imaging
- Capable of identifying 130 different findings
- Attributes confidence bar to suspected findings
- Triage studies based on pre-configured agreed diagnosis list
- Regulated MHRA Class IIb Software as Medical Device

Retrospective analysis suggested high negative predictive value for the tool (>97%) – when it suggested ‘No Findings’ it was usually correct

Deployment:

On the 25th of March 2026 University Hospitals Sussex in collaboration with Harrison.ai and the Surrey, Sussex Frimley Imaging Network went live with an ‘Evaluation Phase’ deployment whereby requesters, clinicians and reporters would have the Annalise.ai CTB inference available to them but with no change to existing human reporting pathways.



Post-Deployment Monitoring

- Human reports compared with AI inference and labelled via PACS interface as True Positive/Negative versus False Positive/Negative
- Combination of reviewers across sites including reporting radiographers, radiologists, specialist neuroradiologists and an enthusiastic Clinical AI Fellow
- Audit period March 25th – End of May (but varied dependent on site)

Results:

	Retrospective	Eastern	Western
N =	2071	2015	1773
Accuracy	78.4%	75.0%	78.1%
Sensitivity (Recall)	89.3%	68.3%	91.7%
Specificity	76.6%	77.9%	76.0%
PPV (Precision)	31.4%	53.6%	37.6%
NPV	97.7%	86.8%	98.3%
False Neg Rate	10.7%	31.6%	8.3%
False Pos Rate	23.4%	22%	24%

Western Site Sub-Analysis

- In a sub-analysis of Western site data - 7 studies labelled as ‘No Findings’ by the AI tool were subsequently reported as having Critical or Urgent findings
- On review, 4 of these were likely human over-calls
- 1 case reported an ocular injury which would be out of scope for the AI tool
- 1 case suggested white matter changes which on subsequent MRI imaging was suggested to be in keeping with demyelinating disease

Radiology trainee feedback

“How has the AI influenced your confidence when calling ‘negative’ scans?”



“Do you find that the tool...?”



Discussion:

Consistent NPV between Retrospective and Western site performance data supports the intended use-case, however, differences between Eastern and Western site performance need further exploration. NPV of 87% potentially unacceptable. High false negative rate risks alert fatigue and disengagement. Why the difference?, possibly:

- Case-mix differences between tertiary trauma/neurosurgical centre - more complex cases seen in Eastern sites
- Higher rates of trauma/skull fracture/facial bone injury or ocular injury in Eastern sites which are known weaknesses of the tool
- Differing skill-mix in reporting personnel (more experienced neuroradiologists in Eastern sites versus general radiologists in Western)
- Different scanners meaning differing image capture

Next Steps:

- Stakeholder meetings to determine ongoing feasibility
- Compare findings across imaging network
- Consider ‘switching off’ findings to improve False neg rate