From pilot to roll-out: a real-world model for AI vigilance and continuous monitoring in fracture detection

Lewisham and Greenwich NHS Trust

Fox Aoife, Hughes Helen, Strappelli Gian-Marco, Clapp Nicola, Thorndyke Steve, Lee Graham, Lewis-Towler Kelly, Harding Elaine, David Sarojini

The Royal College of Radiologists

BACKGROUND

- Missed fractures affect 3–10% of ED cases, causing patient harm and delayed care
- Radiology reporting delays at our trust created a gap between image acquisition and formal diagnosis
- AI-assisted fracture detection was deployed as a point-of-care decision-support tool
- Evaluation and deployment were structured using BS30440:2023, applied prospectively – not as a retrospective compliance exercise,

DEPLOYMENT VOLUME

7,620
Pilot studies
Apr–Aug 2025

30,394
Roll-out studies
Aug 2025–Mar 2026

25,590
Adult studies

4,804
Paediatric studies

Fracture detection rate: 22.5% adults · 24.9% paediatrics

HEALTH ECONOMICS

	Saved (£)
Missed fractures (annual)	**£3,942**
Return to ED (annual)	**£40,976**
Radiologist time (annual)	**£5,840**

Total estimated annual saving **£50,758**

Preliminary analysis. Litigation costs excluded. Causal attribution cannot be confirmed.

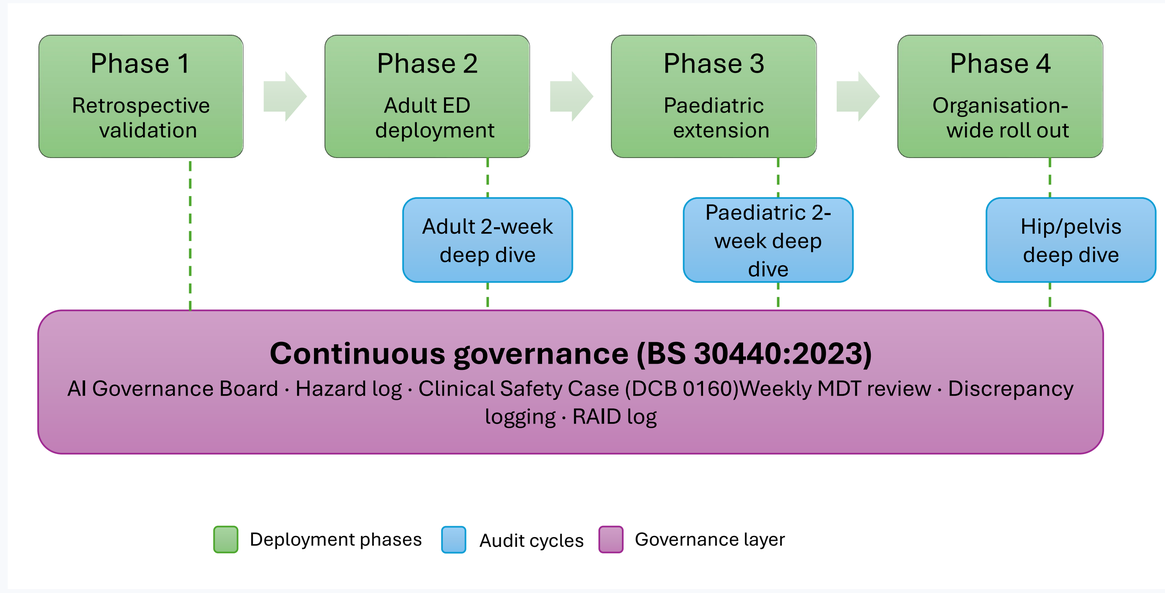
CHALLENGES AND MITIGATIONS

Challenge identified	Action taken
Hip specificity lower than benchmarks (low-confidence findings)	Training · Threshold recalibration
Age-related specificity decline in elderly female patients	Prospective subgroup monitoring
Multi-view inconsistency risk	Mandatory multi-view review · Software update
Periprosthetic fractures missed (n=2)	Senior review mandated · Future: variant training
AI alert not acted upon (3 cases)	Human factors training reinforced
Discrepancy reporting gap	Targeted refresher training
Turnaround concerns at 3 months	Escalated to MDT and resolved with vendor
Dislocation detection misclassification risk	Disabled pre-deployment

Identify → escalate → act

No single monitoring mechanism captured all signals alone.

PHASED IMPLEMENTATION



AI OUTPUT

AI fracture tool output on a paediatric wrist fracture

DIAGNOSTIC PERFORMANCE

ADULT (n= 1058)

99% Sensitivity

94% Specificity

99.3% NPV

89.7% PPV

Exceeded manufacturer benchmarks (93% sensitivity, 90% specificity). Age-related specificity decline in elderly female patients identified through post-deployment audit — not pre-deployment testing.

PAEDIATRICS (n = 237)

98.1% Sensitivity

97.3% Specificity

97.4% NPV

91.4% PPV

Dedicated paediatric audit conducted prior to extension. Growth plates and paediatric anatomy introduce distinct interpretive challenges — performance monitored prospectively before trust-wide scale-up.

HIP DEEP DIVE (N=203, TWO SITES)

90% Sensitivity

93.1% Specificity

98.2% NPV

91.4% PPV

SITE COMPARISON FOLLOWING MODEL VERSION UPDATE

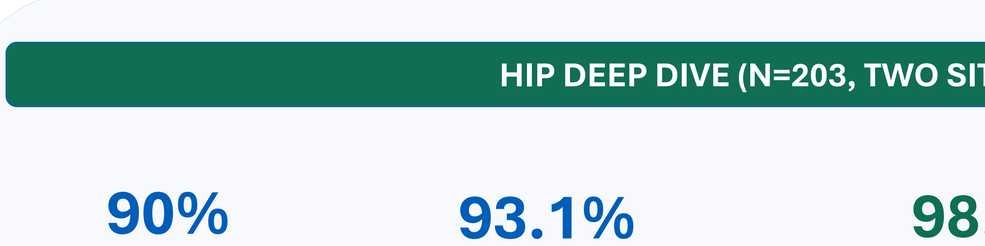
UHL — pre-update (July – August 2025)
94.7 % Sensitivity | 88.1 % Specificity

QEH — post-update (December 2025)
81.8% Sensitivity | 97.8 % Specificity

10% *of patients underwent secondary imaging*

- *85% of secondary imaging cases had discordant AI outputs*
- *Both periprosthetic fractures missed – senior clinical review and radiologist review now implemented*

USER SENTIMENT — ADOPTION TRAJECTORY (n=97)



Trust concerns around overcalling sustained throughout all waves — reflecting healthy critical appraisal, not dissatisfaction.

PATIENT SENTIMENT

(n = 25)

- **47%** comfortable with AI in healthcare
- **50%** agree AI as additional tool makes care safer
- **74%** trust clinicians more than AI

MISSED FRACTURES

20 Pre-deployment July – Sept 2024 → **11** Post-deployment July – Sept 2025

45% reduction

Descriptive comparison only.
3 post-deployment misses had AI alert not acted upon.

RADIOLOGY REPORTING TURNAROUND TIME

13h 15m 2024 mean → **11h 13m** 2025 mean

~2 hr improvement

Observational. Causal attribution to AI deployment cannot be confirmed.

Key conclusions

- High diagnostic performance achieved across adult and paediatric populations in real-world NHS deployment
- Safety signals emerged through structured monitoring — not pre-deployment testing
- Governance framework enabled iterative improvement at model and workflow level
- User adoption: cautious optimism to embedded routine use over 7 months
- Successfully scaled from single-department pilot to trust-wide deployment
- Estimated £50,758 annual cost saving

Good outcomes followed from good process, not good technology alone.