CLINICAL AUDIT

Evaluating the sensitivity and specificity of Gleamer BoneView AI for fracture detection

Dr. L Xu, W. Verrier, Dr. K Skanthabalan, Dr. M Venumbaka

EAST SUSSEX AND NORTH ESSEX NHS FOUNDATION TRUST

East Suffolk and North Essex NHS Foundation Trust

2nd Annual Global AI Conference 2026

-- BACKGROUND

Gleamer BoneView is an AI-powered software designed to assist clinicians in **interpreting trauma musculoskeletal (MSK) radiographs** specifically for the detection of fractures, dislocations, effusions and lesions across various anatomical regions[1]. The software has been in use at our district general hospital (DGH) since June 2024. Literature suggests that BoneView can:

- Enhance diagnostic accuracy
- Increase sensitivity
- Reduce missed fractures[2]

-- AIMS + OBJECTIVES

In this audit, we aimed to measure the **sensitivity** and **specificity** of the Gleamer BoneView software in detecting MSK abnormalities on radiographs. These results helped inform us of the **clinical impact** of BoneView errors.

-- METHODS

STUDY POPULATION

- Retrospective data search on CRIS of adult and paediatric radiographs for MSK trauma, working back from June 2024 to August 2025
- Sample size calculation*

n = 460

EXCLUDED CASES

- No imaging available = 39
- Follow up e.g *post manipulation* = 28
- Out of scope *(age <2y, facial bones, c spine)* = 14
- BoneView not applied = 1

n = 82

COMPLETE CASE SET

- Reporting practitioner compares BoneView interpretation to verified report (AI assisted)
- Patient metrics and TP/TN/FP/FN marked
- Discrepancies reviewed by an arbiter (MSK consultant radiologist) to determine clinical significance of error.
- Statistical analysis: Excel

n = 377

**** The sample size was determined using an appropriate calculator[3] for a single-test design, assuming an expected AI sensitivity of 92%[2] and a disease prevalence of 30%[4-6] .This sample size ensures sufficient statistical power to reliably evaluate AI performance, meaning the study is adequately powered to demonstrate statistically significant results for fracture detection. The assumptions used allow for precise estimation of sensitivity and specificity within an acceptable confidence interval.***

-- RESULTS

377 cases were assessed with a total of **109 fractures** (28.9%). BoneView showed a sensitivity of 92.6% and specificity 91.4% for fracture detection; of these, 9 false negatives and 22 false positives were identified (fig.1). Themes identified within discrepant cases (false positives and false negatives) included presence of old fractures, ossicles, growth plates and artefact or metallic prosthesis (fig 2,3). **Only one discrepant case** (false positive i.e. overcall) was **deemed clinically significant** as it led to unnecessary further imaging (fig 2A). The only other clinically significant BoneView error was of one missed dislocation.

Prevalence of **effusion** (10.1%), **lesion** (1.6%) and **dislocation** (4.3%) were too low to enable any meaningful statistical analysis on BoneView's ability to detect these pathologies.

	FRACTURE PRESENT	FRACTURE ABSENT
BoneView RESULT POSITIVE	TRUE POSITIVE 100	FALSE POSITIVE 22
BoneView RESULT NEGATIVE	FALSE NEGATIVE 9	TRUE NEGATIVE 246

Fig 1. Contingency table

N = 377

92.6% Sensitivity

91.4% Specificity

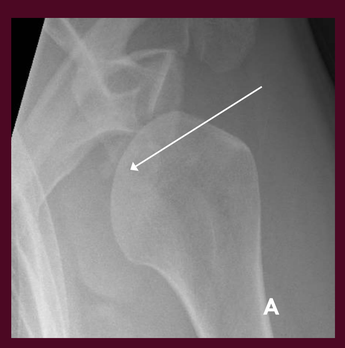

Fig 2. False positive (i.e. under called) case for a Bankart fracture (A). This was a non-significant error as it was picked up on post manipulation imaging. Shoulder dislocation was correctly identified (B).

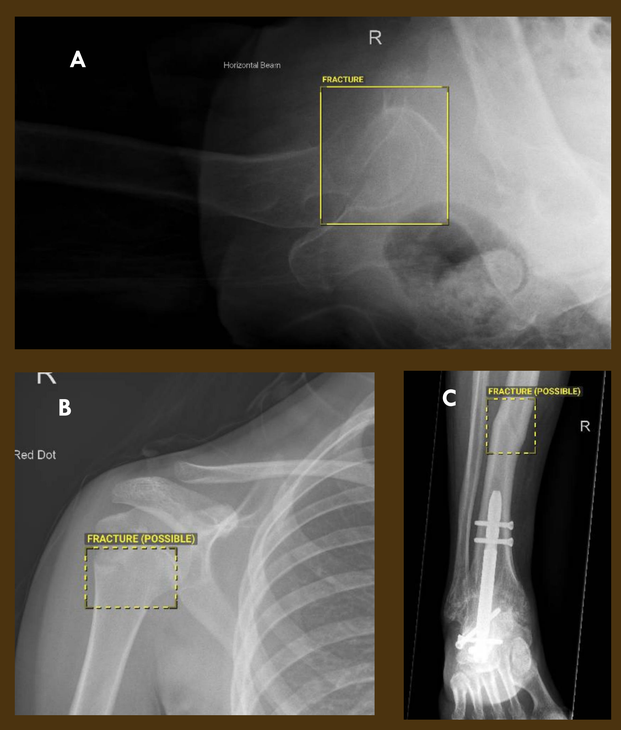

Fig 3. False positive cases. Case A was a significant error as this resulted in a CT hip that was not required. Other common false positives included growth plates (B) and old fractures (C).

-- CONCLUSION

BoneView showed a **high sensitivity and specificity** for fracture detection which was expected, and interestingly better than reported in some previous studies[2]. Overall, it has proven to have a **very low rate of significant clinical error**.

This work supports the notion that AI is a useful software adjunct for reporters to optimise detection of MSK abnormalities and boost workflow.

REFERENCES

1) Gleamer AI https://www.gleamer.ai/ **2) Kuo et al. (2022)** *Radiology*, **304**(3), pp. 512–523. https://doi.org/10.1148/radiol.211785 **3) Akoglu (2022)** *Turkish Journal of Emergency Medicine*, **22**(4). **4) Willis and Sur (2007)** *European Journal of Emergency Medicine*, **14**(1), pp. 14–16. **5) Hardy, Snaith and Scally (2013)** *British Journal of Radiology*, **86**(1021). **6) Brealey (2001)** *Radiography*, **7**(4), pp. 263–270.