Real world comparison between the HealthVCF and HealthOST AI models for detection of vertebral fractures from existing CT scans

Authors: R Eckert[1], M. K. Javaid[2]

[1] Oxfordshire Fracture Prevention Service Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford, OX3 7HE, UK.
[2] International XLH Alliance, London, UK

Background & Objectives

- Reducing the number of adults presenting with hip fractures is a major public health priority.
- The focus has been on implementation of effective secondary fracture prevention using the Fracture Liaison Service (FLS) model to identify and manage adults presenting with a major osteoporotic fracture to hospitals.
- Even though vertebral fractures (VF) strongly predict hip fracture risk, most adults with VFs are missed by FLSs.
- It is established that over 10% of CT scans performed in adults aged 50 years and over, have a VF but are both not often reported and even if reported do not lead to effective secondary fracture prevention.
- A number of different AI models have been developed to identify vertebral fractures (VFs) from existing CT scans to increase the case identification of adults with VFs.

We aimed to compare the test performance of AI detected vertebral fractures (moderate/ severe deformity or clear endplate fractures) between the commercially available Nanox-AI HealthVCF (VCF) and HealthOST (OST) models.

Methods

- A consecutive series of 431 CT scans that included the thoracic or lumbar spine from 1st to 3rd October 2025 were analysed using the VCF and OST models using the balanced sensitivity equivalent settings.
- While VCF analyses axial images to generate a sagittal image output, HealthOST uses the primary sagittal images to measure vertebral height. HealthOST also flags scans with low BMD.
- Both models generate within the On-Premise Interface (OPI) a table with patient demographics and a sagittal image including a box for potential VFs.
- An FLS nurse, with experience in identifying vertebral fractures. read the AI flagged positive and negative CT scans for both VCF and OST blinded to order to identify VF with moderate/ severe deformity or a clear endplate fracture.
- The AI flag rate, sensitivity, specificity, negative and positive predictive value (NPV, PPV) were calculated as well as the estimated impact per 1000 scans using a 12.7% VF prevalence rate.
- Descriptive statistics were used to test agreement (Cohen's kappa, Gwets AC) and describe differences (Fishers exact test) between the two AI models.

Results

- Of the consecutive scans, 364 were analysed by both models and included in the results.
- The clinician confirmed VF prevalence was 9.6% with VCF and significantly higher at 12.7% with OST ($p<0.001$) with 94.8% agreement (kappa 0.74,; Gwet's AC 0.93, $p<0.001$).

Performance characteristics from VCF to OST including estimated impact per 1000 scans

	AI flag rate	Sensitivity	Specificity	% Clinician agreement	NPV	PPV	Estimated impact per 1000 scans using a 12.7% VF prevalence			
							AI flagged scans (n)	AI flagged with VF (n)	AI flagged without VF (n)	VF missed
HealthVCF	32.7%	88.6%	73.3%	74.7%	98.4%	26.1%	327	112	215	14.5
HealthOST	28.0%	95.7%	81.7%	83.5%	99.2%	43.1%	280	121	159	5.4

- 10.1% of OST AI scans were also identified as potentially low BMD
- There where significantly higher rates of low BMD in patients that also had clinically confirmed VFs (23.9% vs 8.2%, $p=0.001$)

Strengths and limitations

- This was a real-world comparison of 2 AI models
- The VF were confirmed by a trained FLS nurse not musculoskeletal radiologist as standard of care.

Conclusion

- The updated OST model improved both clinical confirmation of VF as well as model performance in a real-world FLS setting, flagging fewer scans for clinical confirmation and missing significantly fewer VFs.
- Using trained FLS nurses to confirm VF represents a scalable option where radiologist access is limited.
- The integration of the AI low BMD status on FLS triage and management of patients requires further follow-up.

Kassim.Javaid@ndorms.ox.ac.uk