Osteoporosis International (2025) 36:1053–1060
https://doi.org/10.1007/s00198-025-07487-0

ORIGINAL ARTICLE

Opportunistic assessment of osteoporosis using hip and pelvic X-rays with OsteoSight™: validation of an AI-based tool in a US population

Robert J. Pignolo[1] · John J. Connell[2] · Will Briggs[2] · Catherine J. Kelly[2] · Chris Tromans[2] · Naima Sultana[2] · J. Michael Brady[2]

Received: 4 February 2025 / Accepted: 28 March 2025 / Published online: 22 April 2025

Abstract

Summary Identifying patients at risk of low bone mineral density (BMD) from X-rays presents an attractive approach to increase case finding. This paper showed the diagnostic accuracy, reproducibility, and robustness of a new technology: OsteoSight™. OsteoSight could increase diagnosis and preventive treatment rates for patients with low BMD.

Purpose This study aimed to evaluate the diagnostic accuracy, reproducibility, and robustness of OsteoSight™, an automated image analysis tool designed to identify low bone mineral density (BMD) from routine hip and pelvic X-rays. Given the global rise in osteoporosis-related fractures and the limitations of current diagnostic paradigms, OsteoSight offers a scalable solution that integrates into existing clinical workflows.

Methods Performance of the technology was tested across three key areas: (1) diagnostic accuracy in identifying low BMD as compared to dual-energy X-ray absorptiometry (DXA), the clinical gold standard; (2) reproducibility, through analysis of two images from the same patient; and (3) robustness, by evaluating the tool's performance across different patient demographics and X-ray scanner hardware.

Results The diagnostic accuracy of OsteoSight for identifying patients at risk of low BMD was area under the receiver operating characteristic curve (AUROC) 0.834 [0.789–0.880], with consistent results across subgroups of clinical confounders and X-ray scanner hardware. Specificity 0.852 [0.783–0.930] and sensitivity 0.628 [0.538–0.743] met pre-specified acceptance criteria. The pre-processing pipeline successfully excluded unsuitable cases including incorrect body parts, metalwork, and unacceptable femur positioning.

Conclusion The results demonstrate that OsteoSight is accurate in identifying patients with low BMD. This suggests its utility as an opportunistic assessment tool, especially in settings where DXA accessibility is limited or not recently performed. The tool's reproducibility and robust performance across various clinical confounders further supports its integration into routine orthopedic and medical practices, potentially broadening the reach of osteoporosis assessment and enabling earlier intervention for at-risk patients.

Keywords BMD · Case-finding · Osteopenia · Osteoporosis · X-rays

Introduction

Osteoporosis is characterized by the progressive loss of bone mineral density (BMD) which increases the risk for fragility and other fractures. One in five men and one in three women over the age of 50 will experience an osteoporotic fracture, resulting in loss of mobility, reduced quality of life, and increased mortality [1]. Annual direct health costs of osteoporotic fractures are estimated at $57 billion in 2018 in the USA [2]. It is estimated that fewer than 26% of people living with osteoporosis have been diagnosed [3], despite strong evidence that treatment can prevent fragility fractures and maintain an individual's quality of life [4]. Identification of those individuals at risk of osteoporosis is therefore a significant public health concern.

Osteoporosis diagnosis is typically supported using dual-energy X-ray absorptiometry (DXA), which is considered

✉ John J. Connell
john.connell@naitive.com

[1] Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA

[2] Naitive Technologies Ltd, London EC1N 2SW, UK

the reference standard test. The US Preventive Services Task Force recommends screening for osteoporosis with DXA in women of 65 years and over or younger women with certain clinical risk factors; insufficient evidence is available for recommendations in men [5, 6]. Additional support for screening has been voiced by the American Association of Clinical Endocrinologists [7], Bone Health and Osteoporosis Foundation [8], American College of Obstetrics & Gynecology [9], and the Osteoporosis International clinician's guide [10]. Prioritizing preventive care using screening for a "silent disease" is challenging. In a recent survey on barriers to osteoporosis screening in the USA, physicians reported visit time constraints and low prioritization against competing health concerns and a lack of awareness and/or timely access to services and resources as key barriers to guideline adherence for osteoporosis screening [11, 12]. Though the benefits of screening are evident, screening rates are low, with attendance at targeted programs as low as 1.7% for men and 9.5% for women [13]. Case-finding of the many individuals with low BMD, but who are not aware of the fact, is a growing global unmet medical need.

One approach to addressing this is through the opportunistic assessment for signs of osteoporosis on X-rays that are being collected for other indications. Advances in automated image analysis and machine learning have made it possible to evaluate patients at scale, cost-effectively, for multiple indications [14]. These include breast cancer [15], cardiovascular disease [16], and increasingly lung cancer.

OsteoSight™ is a machine learning-based technology that processes hip or pelvic digital X-ray images at scale. OsteoSight receives images from a healthcare institution's picture archiving and communication system (PACS) and assesses X-ray metadata for eligibility (aged 50 years and over, AP X-rays of the hip or pelvis). OsteoSight performs a series of image-based quality control checks to ensure that images with prostheses or incorrect body parts (e.g., chest X-ray) are not used to generate a result. An estimated BMD is derived from the proximal femur, and if the patient is at risk of low BMD, this is flagged to the physician interpreting the X-ray, within their standard clinical workflow, who may then include a recommendation for a bone health assessment and DXA scan in their findings.

Gaining clinical acceptance of such a device requires rigorous evaluation of the results compared to the current clinical gold standard, as well as evidence that assures robustness of results on repeat tests and specific X-ray scanner hardware. Concerns regarding generalizability of machine learning-based algorithms are considerable, with regulatory agencies requiring that manufacturers of such technologies demonstrate performance in independent test populations that are representative of the intended use population. It is important that these test populations are as realistic as possible and include common challenges encountered in the real world, such as cases of mislabeled body parts, suboptimal patient positioning, and the presence of prostheses.

In this study, we evaluate the accuracy, reproducibility, and robustness of OsteoSight, based on a DXA diagnosis of low BMD, in an independent non-curated dataset from community imaging centers offering X-ray services.

Materials and methods

A series of performance tests of OsteoSight were performed to evaluate: (i) the diagnostic accuracy of the device for providing a flag of low BMD; [2] the reproducibility of the device to provide the same results on repeat tests of the same patient; and [3] the impact of clinical confounders on the diagnostic accuracy.

Study population

Data from a medical imaging broker, Gradient Health, was collected from six US imaging centers and made available for ethical research. Data was de-identified, complying with HIPAA's safe harbor protocol; as the data is de-identified and poses minimal risk to privacy and confidentiality, IRB review was not deemed necessary [17]. Pre-specified inclusion criteria based on the image metadata were applied to data curation protocols: female and male individuals aged 50 years and over; a hip or pelvic anteroposterior (AP) X-ray image; a DXA scan collected within ± 12 months of the X-ray; and minimal required demographic information including age at time of first scan, ethnicity, and BMI.

Reference standard—DXA

The clinical reference standard for diagnosis of low BMD is a T-score of less than − 1.0 standard deviations (SD) for osteopenia and − 2.5 SD for osteoporosis, as measured with DXA. In this study, individuals were classified as having low BMD if they had a T-score less than − 1.0 SD at the femoral neck (FN) as OsteoSight estimates BMD specifically at this anatomical site. While osteoporosis diagnosis typically considers the lowest T-score from multiple skeletal sites (e.g., FN, total hip, or lumbar spine), we selected the FN T-score for classification to ensure a direct comparison between OsteoSight's eBMD estimates and DXA-derived BMD at the same location. T-scores were derived from BMD measurements by the DXA manufacturer on the de-identified DXA report. All BMD values in this study were measured using Hologic densitometers.

X-ray pre-processing and BMD inference

OsteoSight comprises a sequence of image analysis steps to ensure that only appropriate X-ray images are used to estimate BMD, as the most common reasons for rejected radiographs in clinical practice are due to mislabeling (22%) and positioning (16%) [18]. First, clinically unsuitable images are excluded, specifically in cases when metalwork (e.g., implants and clothing accessories) is detected, or the wrong body part is presented. Next, the algorithm attempts to identify a suitable proximal femur region of interest for analysis. An example AP hip X-ray that successfully passed the pre-processing steps (Fig. 1) is shown, with the (green) proximal femur segmentation indicated including the femoral neck, greater trochanter, and the intertrochanteric region, similar to region of interest methods used in DXA [19]. If the femur is not fully contained within the field of view, or it cannot be segmented accurately (e.g. due to suboptimal image acquisition and gross anatomical morphologies), the image is excluded.

Then the image is linearly normalized in preparation for processing using a convolutional neural network (CNN) which has been trained using X-ray images paired with corresponding DXA results using a fivefold cross-validation method to generate an estimated BMD (eBMD) value. The anonymized data used to train the CNN were acquired between 2011 and 2021 from the Royal National Orthopedic Hospital NHS Trust, UK. During training and pre-training, it was ensured that no images from subjects in the test set were in the training or validation sets.

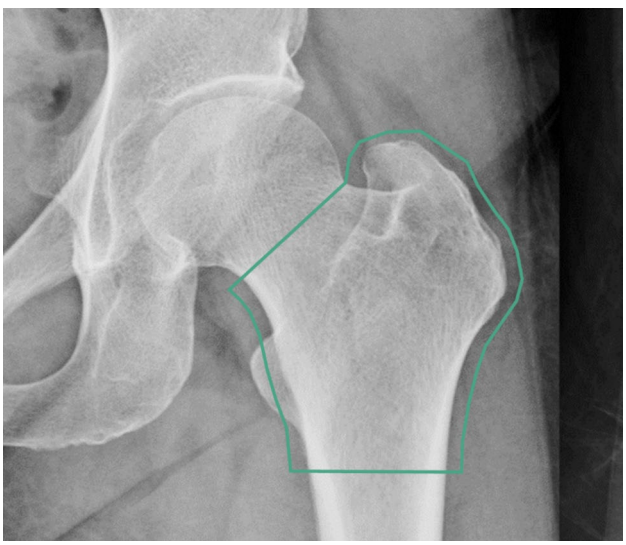

Fig. 1 Example AP hip X-ray image showing region of interest (green) analyzed to estimate BMD following OsteoSight's multi-step image quality and verification processing pipeline

Statistical analyses

All statistical analyses were performed in R version 4.3.2 (Posit Software). For diagnostic accuracy studies, the eBMD was compared to the presence of low BMD as measured by DXA (femoral neck T-score < -1.0 SD) and area under the receiver operating characteristic curve (AUROC) were computed, with 95% confidence intervals (CI) with a primary endpoint pre-specified acceptance criteria of AUROC (lower CI > 0.75). To evaluate the performance of OsteoSight, sensitivity and specificity were evaluated across the study population with pre-specified acceptance criteria of specificity (lower CI > 0.775) and sensitivity (lower CI > 0.5) [20].

The 95% CI for the AUROC, sensitivity, and specificity were calculated with 2000 stratified bootstrap replicates using the pROC library [21]. A series of diagnostic accuracy subgroup analyses were conducted, as above, to evaluate the robustness of the results when considering the most relevant clinical confounders of age, sex, ethnicity, BMI, and X-ray scanner hardware manufacturer.

To evaluate the reproducibility of OsteoSight, a subset of individuals were identified where two X-rays of the same hip had been collected. The agreement of the OsteoSight flag for the presence of low BMD was assessed using Cohen's Kappa test.

Sample size

The minimum sample size required to power the primary endpoint was calculated [22] as 225 subjects, assuming power $= 80\%$, alpha $= 0.025$, estimating disease prevalence $= 43.1\%$, and confidence interval full width $= 0.10$.

Results

The performance of OsteoSight in removing unsuitable X-ray images is outlined in Fig. 2. From an initial population of 378 subjects, 22 cases were excluded due to a mislabeled body part (i.e., not a hip or pelvic radiograph), 0 cases were excluded due to the hip containing metalwork in the region of interest (ROI), and 31 cases were excluded due to their images not containing a suitable femoral ROI (e.g., proximal femur extends beyond field of view, anatomical rotation as in the case of a frog leg lateral image, or gross anatomical deformity due to a displaced fracture). Excluded and included images were confirmed by manual inspection. BMD estimation was then successfully completed for all remaining cases successfully completed, and 325 cases (86%) were included in the study population for subsequent analyses to represent one case per patient, by selecting the image closest in date to the reference DXA report. The demographics of all individuals within the study population

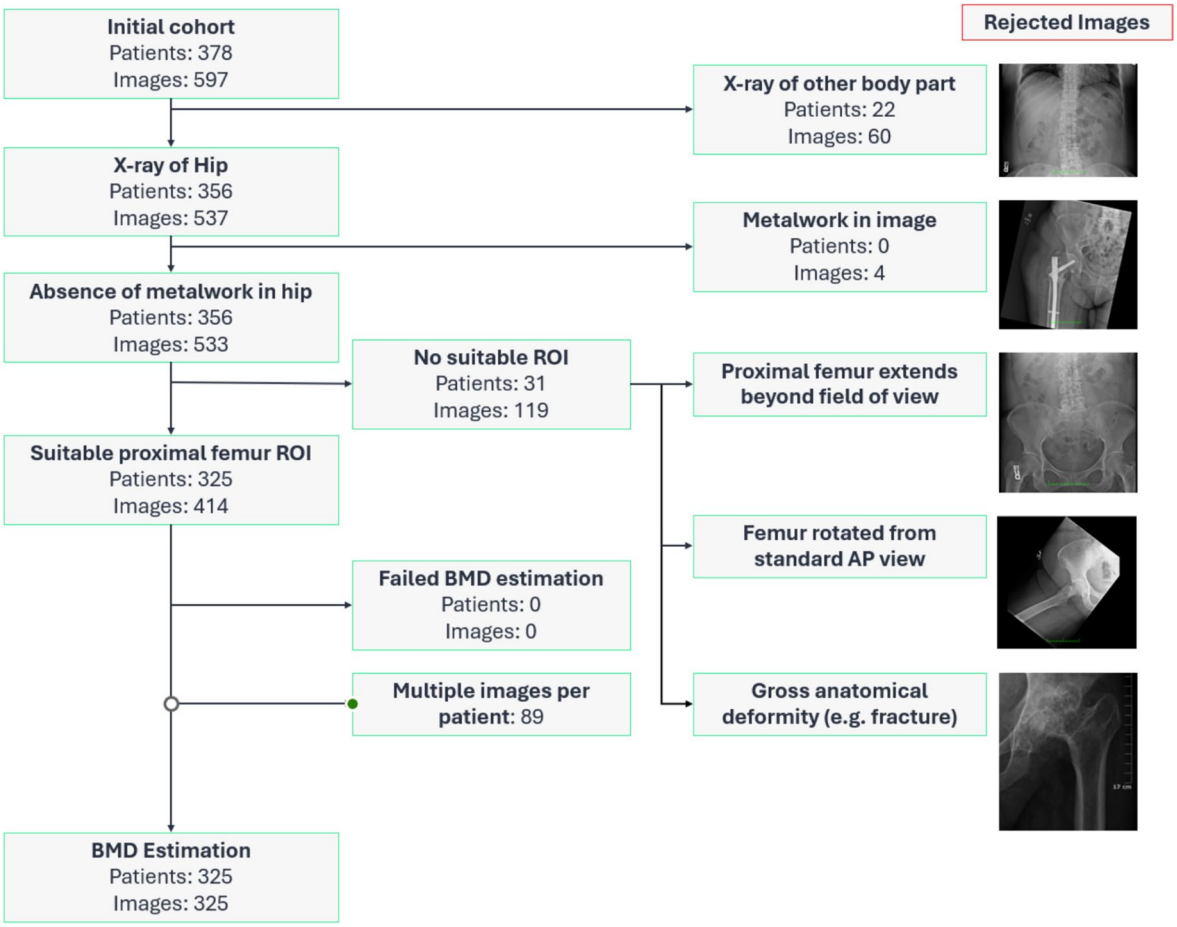

Fig. 2 Consort diagram detailing OsteoSight's multi-step image quality and verification processing pipeline. Region of interest (ROI), anteroposterior (AP), and bone mineral density (BMD)

Table 1 Study population demographics following the image quality and verification processing pipeline. Standard deviation (SD); bone mineral density measured using dual energy X-ray absorptiometry (BMD)

	Study population
N	325
Age (years, mean [SD])	69.9 [10]
Sex female: male	310: 15
BMI (kg/m^2, mean [SD])	32.5 [8.46]
Race	Hispanic 141 White 134 Black or African American 36 Asian 14
X-ray scanner	Konica Minolta 301 Siemens 24
Femoral neck BMD (g/cm^2, mean [SD])	0.708 [0.144]

are reported in Table 1. 64.6% of the study population had a femoral neck T-score measured by DXA as lower than − 1.0 SD.

The diagnostic accuracy of OsteoSight for identifying patients at risk of low BMD was AUROC 0.834 [0.789–0.880] (Fig. 3). At a model operating point (Fig. 3, red point) selected to satisfy both pre-specified performance acceptance criteria, the specificity was 0.852 [0.783–0.930], and the sensitivity was 0.628 [0.538–0.743] indicating that the primary and secondary endpoints of this study were met (Table 2). A confusion matrix (Table 3) indicates that this model operating point minimizes false positives (17/325 or 5%).

Subset analysis of clinically relevant confounders showed a diagnostic accuracy of OsteoSight consistent across age groups, sex, ethnicity, BMI, and all tested X-ray scanner hardware (Table 4). The lower bound confidence interval for individuals in the Siemens group, 18.5–25 BMI group, and 70–79 and 80 + age groups were below the pre-specified acceptance criteria of 0.75.

A subset of 81 individuals where two acceptable hip or pelvic X-ray images of the same anatomical laterality analyzed by OsteoSight showed an 84% observed agreement, resulting in a Cohen's kappa of 0.68 (Table 5) reflecting

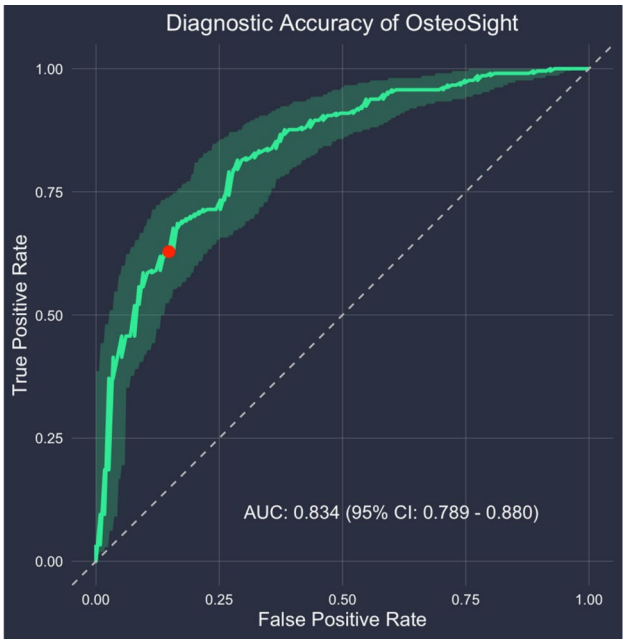

Fig. 3 Receiver operating characteristic (ROC) curve showing diagnostic accuracy of OsteoSight for predicting individuals with low bone mineral density. Red dot indicates preferred operating threshold; shaded region indicates sensitivity confidence intervals (CI). Area under the curve (AUC)

Table 2 Performance of OsteoSight at a single operating threshold. [95% confidence intervals]

	Specificity	Sensitivity
OsteoSight performance	0.852 [0.783–0.930]	0.628 [0.538–0.743]

Table 3 Confusion matrix for performance of OsteoSight

		Actual	
Predicted		Negative	Positive
	Negative	98	78
	Positive	17	132

substantial agreement [23]. There was an average of 128 days between repeat X-rays.

Discussion

The results shown here indicate that OsteoSight can identify individuals who are at risk of low BMD from routine X-rays, with a high level of diagnostic accuracy, reproducibility, and robustness. The sensitivity and specificity of OsteoSight support its use as an opportunistic assessment tool to detect low BMD with performance metrics that align with those of DXA [24, 25], and the diagnostic accuracy is equivalent to systems that have previously been presented in the literature using X-rays or CT scans [20, 26]. This is a promising development, addressing at least some of the challenges associated with DXA. By enabling the detection of low BMD from X-rays, generally taken for some other clinical reason, OsteoSight could significantly increase osteoporosis case finding, particularly in resource-limited settings where DXA access is unavailable, limited, or otherwise not performed.

We have also shown that the technology automatically identifies and filters out incorrectly labeled X-ray images as well as a number of artifacts and unsuitable images, aiming to render the technology as a more reliable and explainable AI solution. Many imaging technologies do not perform as well as anticipated in real-world testing [27]. The technology assessed here begins to address this issue as the test data was sourced from hospitals' PACS systems, where mislabeling could happen, rather than from collected and curated research datasets.

The reproducibility of OsteoSight was shown by the consistent results for multiple X-ray images taken at different times from the same individuals. The 84% observed agreement aligns with BMD mis-classification estimates from DXA of 10–50% [28]. This reproducibility is critical for clinical adoption, ensuring that the tool's output is dependable and that patients flagged by the system truly require further evaluation.

Finally, the robustness of OsteoSight was tested across different patient subgroups and clinical confounders and across a range X-ray scanner hardware. This study population was representative of the eventual target patient population, despite some subgroups being underpowered. Individuals aged 50 and over were included in this study as this is the youngest age at which individuals are considered for osteoporosis screening in the USA [5]. The consistency of results supports the applicability of OsteoSight deployment in a wide range of clinical environments.

Clinical implications

The majority of hip X-rays are ordered by orthopedic surgeons [29], and they are generally interpreted either by radiologists or orthopedic surgeons depending on the healthcare setting [30]. However, the results are also relevant to primary care, geriatrics, endocrinology, gynecology/women's health, and rheumatology. OsteoSight can be deployed in these diverse clinical settings enabling clinicians to (i) maximize specificity (false positives) and the associated risks and costs of subsequent follow-up patient management while retaining clinically acceptable sensitivity and/or (ii) minimize the number of missed cases by maximizing sensitivity while retaining clinically acceptable specificity.

Table 4 AUC values of subgroups showing diagnostic accuracy of OsteoSight for predicting individuals with low bone mineral density. Area under the receiver operating characteristic curve, AUC; confidence intervals, CI. Six individuals omitted with BMI < 18.5 kg/m^2; seven individuals omitted with missing BMI data

Group	Subgroup	AUC	Lower 95% CI	Upper 95% CI	Sample size
Race	Hispanic	0.820	0.750	0.891	141
Race	White	0.829	0.756	0.902	134
Race	Black or African American	0.892	0.771	1.000	36
Race	Asian	1.000	1.000	1.000	14
Sex	Female	0.824	0.776	0.872	310
Sex	Male	1.000	1.000	1.000	15
X-ray scanner	Konica	0.833	0.785	0.881	301
X-ray scanner	Siemens	0.829	0.660	0.997	24
Age (years)	50–59	0.810	0.693	0.928	54
Age (years)	60–69	0.865	0.795	0.936	104
Age (years)	70–79	0.788	0.693	0.882	104
Age (years)	80 +	0.778	0.624	0.933	61
BMI (kg/m^2)	$18.5 \leq 25$	0.763	0.525	1.000	57
BMI (kg/m^2)	$25 \leq 30$	0.853	0.769	0.938	86
BMI (kg/m^2)	30 +	0.811	0.746	0.876	169

Table 5 Agreement of repeat X-ray analyses by OsteoSight

Metric	Value
Observed agreement	84% (68/81)
Cohen's kappa	0.68

The growing global burden of osteoporosis-related fractures highlights the need for a scalable, cost-effective, and efficient alternative method that can be integrated seamlessly into existing clinical workflows and infrastructure. The integration of OsteoSight into routine clinical practice and using X-rays collected for other indications (e.g., hip pain and rule out fracture) could facilitate earlier identification of patients at risk for osteoporosis, thereby enabling timely intervention and potentially reducing the incidence of osteoporotic fractures.

The missed opportunity to get the right patients onto the right treatment pathways is a growing healthcare and economic problem. The chronically poor rates of adherence to screening programs have resulted in millions of individuals receiving a diagnosis only after a fragility fracture has occurred. The use of OsteoSight, as an opportunistic assessment tool to evaluate hip and pelvic X-rays that are collected for other indications, has the potential to significantly improve the case finding of osteoporosis and osteopenia. Its successful integration into clinical practice could play a crucial role in addressing the growing burden of osteoporosis by enabling earlier detection and intervention, ultimately improving patient outcomes and reducing healthcare costs.

Limitations and future research

While this study included a diverse patient population sourced from multiple independent sites, it remains a single validation dataset. An increased sample size, particularly in the subgroups, would further reassure that the results are generalizable to broader clinical settings. The training dataset was sourced from an orthopedic hospital and may disproportionately reflect patients undergoing an X-ray for osteoarthritis. Also, the study did not assess the long-term outcomes of patients, which are crucial for understanding the impact on patient management and decisions around providing treatment options.

Future research should focus on prospective studies to validate the findings presented here in real-world clinical settings: assessing the impact of OsteoSight on patient outcomes, including fracture rates and quality of life; and exploring integration of OsteoSight with clinical decision support systems to further enhance its utility in osteoporosis management. Regardless of model performance, the success of tools will be dependent on their implementation within the broader care pathway in collaboration with clinicians. Ensuring that no patients are lost, from identification, through diagnosis, to therapy, is critical.

Conclusions

This study demonstrates the potential utility of OsteoSight as a clinical tool for early case finding of individuals at risk of low BMD, using routine hip and pelvic X-rays. The machine learning-based algorithm which underpins

OsteoSight has high diagnostic accuracy when compared to the ground truth and shows generalizability across clinical confounders.

Data availability Data available upon reasonable request.

Declarations

Conflict of interest John Connell, Will Briggs, Catherine Kelly, Chris Tromans, and Naima Sultana are employees and share (option) holders at Naitive Technologies Ltd. Mike Brady is a non-executive director and share (option) holder at Naitive. Robert J. Pignolo is a member of the Naitive scientific advisory board. Ethics approval was not necessary for this study. Naitive Technologies Ltd owns patents: PCT/IB2024/050708 and PCT/IB2024/056808, which are related to the technology/methodology described in this study.

References

1. (n.d) Epidemiology | International Osteoporosis Foundation [Internet]. Available from: https://www.osteoporosis.foundation/health-professionals/fragility-fractures/epidemiology#ref_bottom_4. Accessed 14 Nov 2024
2. Lewiecki EM, Ortendahl JD, Vanderpuye-Orgle J, Grauer A, Arellano J, Lemay J et al (2019) Healthcare policy changes in osteoporosis can improve outcomes and reduce costs in the United States. JBMR Plus 3(9):e10192. https://doi.org/10.1002/jbm4.10192
3. Weaver J, Sajjan S, Lewiecki EM, Harris ST (2017) Diagnosis and treatment of osteoporosis before and after fracture: a side-by-side analysis of commercially insured and medicare advantage osteoporosis patients. J Manag Care Spec Pharm 23(7):https://doi.org/10.18553/jmcp.2017.23.7.735
4. Makridis KG, Karachalios T, Kontogeorgakos VA, Badras LS, Malizos KN (2015) The effect of osteoporotic treatment on the functional outcome, re-fracture rate, quality of life and mortality in patients with hip fractures: a prospective functional and clinical outcome study on 520 patients. Injury 46(2):378–383
5. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB et al (2018) Screening for osteoporosis to prevent fractures: US Preventive Services Task Force recommendation statement. JAMA 319(24):2521–31
6. (n.d) Screening to prevent osteoporotic fractures: updated evidence report and systematic review for the US Preventive Services Task Force | Guidelines | JAMA | JAMA Network [Internet]. https://doi.org/10.1001/jama.2018.6537
7. Camacho PM, Petak SM, Binkley N, Diab DL, Eldeiry LS, Farooki A et al (2020) American Association of Clinical Endocrinologists/American College of Endocrinology clinical practice guidelines for the diagnosis and treatment of postmenopausal osteoporosis—2020 update. Endocr Pract 1(26):1–46
8. Cosman F, de Beur SJ, LeBoff MS, Lewiecki EM, Tanner B, Randall S et al (2014) Clinician's guide to prevention and treatment of osteoporosis. Osteoporos Int 25(10):2359–2381
9. White L (2022) Osteoporosis prevention, screening, and diagnosis: ACOG recommendations. Am Fam Physician 106(5):587–588
10. LeBoff MS, Greenspan SL, Insogna KL, Lewiecki EM, Saag KG, Singer AJ et al (2022) The clinician's guide to prevention and treatment of osteoporosis. Osteoporos Int 33(10):2049–2102
11. Bennett MJ, Center JR, Perry L (2023) Exploring barriers and opportunities to improve osteoporosis care across the acute-to-primary care interface: a qualitative study. Osteoporos Int 34(7):1249
12. Sabri SA, Chavarria JC, Ackert-Bicknell C, Swanson C, Burger E (2023) Osteoporosis: an update on screening, diagnosis, evaluation, and treatment. Orthopedics 46(1):e20–e26
13. Pisu M, Kopperdahl DL, Lewis CE, Saag KG, Keaveny TM (2019) Cost-effectiveness of osteoporosis screening using biomechanical computed tomography for patients with a previous abdominal CT. J Bone Miner Res Off J Am Soc Bone Miner Res 34(7):1229–1239
14. Söreskog E, Lopez B, Bean T, Lewis P, Ashley N, Lopes JDP et al (2024) Cost-effectiveness and societal burden implications of screening for fracture risk in a UK general radiography setting [Internet]. Available from: https://www.researchsquare.com/article/rs-4739580/v1. Accessed 6 Sept 2024
15. Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H et al (2023) Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol 24(8):936–944
16. Patel MR, Nørgaard BL, Fairbairn TA, Nieman K, Akasaka T, Berman DS et al (2020) 1-year impact on medical practice and clinical outcomes of FFRCT: the ADVANCE registry. JACC Cardiovasc Imaging 13(1 Pt 1):97–105
17. (n.d) HIPAA privacy rule and its impacts on research [Internet]. Available from: https://privacyruleandresearch.nih.gov/irbandprivacyrule.asp. Accessd 14 Nov 2024
18. Alyousef KA, Alkahtani S, Alessa R, Alruweili H (2020) Radiograph reject analysis in a large tertiary care hospital in Riyadh, Saudi Arabia. Glob J Qual Saf Healthc 2(2):30–33
19. Lorente-Ramos R, Azpeitia-Armán J, Muñoz-Hernández A, García-Gómez JM, Díez-Martínez P, Grande-Bárez M (2011) Dual-energy X-ray absorptiometry in the diagnosis of osteoporosis: a practical guide. Am J Roentgenol 196(4):897–904
20. Bilbily A, Syme CA, Adachi JD, Berger C, Morin SN, Goltzman D et al (2024) Opportunistic screening of low bone mineral density from standard X-rays. J Am Coll Radiol 21(4):633–639
21. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 17(12):77
22. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36
23. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174
24. Williams S, Khan L, Licata AA (2021) DXA and clinical challenges of fracture risk assessment in primary care. Cleve Clin J Med 88(11):615–622
25. Marques A, Ferreira RJO, Santos E, Loza E, Carmona L, da Silva JAP (2015) The accuracy of osteoporotic fracture risk prediction

tools: a systematic review and meta-analysis. Ann Rheum Dis 74(11):1958–1967
26. Yen TY, Ho CS, Chen YP, Pei YC (2024) Diagnostic accuracy of deep learning for the prediction of osteoporosis using plain X-rays: a systematic review and meta-analysis. Diagnostics 14(2):207
27. Smets J, Shevroja E, Hügle T, Leslie WD, Hans D (2021) Machine learning solutions for osteoporosis—a review. J Bone Miner Res 36(5):833–851
28. Phillipov G, Seaborn CJ, Phillips PJ (2001) Reproducibility of DXA: potential impact on serial measurements and misclassification of osteoporosis. Osteoporos Int 12(1):49–54
29. Harkey P, Duszak R, Gyftopoulos S, Rosenkrantz AB (2018) Who refers musculoskeletal extremity imaging examinations to radiologists? Am J Roentgenol 210(4):834–841
30. Magnuson JA, Parikh N, Sirch F, Montgomery JR, Kyriakos RN, Saxena A et al (2024) Is the interpretation of radiographic knee arthritis consistent between orthopaedic surgeons and radiologists? J Orthop Exp Innov [Internet]. 5(1). https://doi.org/10.60118/001c.91022