



■ HIP

Development and validation of an open-source tool for opportunistic screening of osteoporosis from hip CT images

A MULTICENTRE STUDY OF 978 HIPS

**K. Uemura,
Y. Otake,
K. Takashima,
H. Hamada,
T. Imagama,
M. Takao,
T. Sakai,
Y. Sato,
S. Okada,
N. Sugano**

From Osaka University,
Suita, Japan

Aims

This study aimed to develop and validate a fully automated system that quantifies proximal femoral bone mineral density (BMD) from CT images.

Methods

The study analyzed 978 pairs of hip CT and dual-energy X-ray absorptiometry (DXA) measurements of the proximal femur (DXA-BMD) collected from three institutions. From the CT images, the femur and a calibration phantom were automatically segmented using previously trained deep-learning models. The Hounsfield units of each voxel were converted into density (mg/cm^3). Then, a deep-learning model trained by manual landmark selection of 315 cases was developed to select the landmarks at the proximal femur to rotate the CT volume to the neutral position. Finally, the CT volume of the femur was projected onto the coronal plane, and the areal BMD of the proximal femur (CT-aBMD) was quantified. CT-aBMD correlated to DXA-BMD, and a receiver operating characteristic (ROC) analysis quantified the accuracy in diagnosing osteoporosis.

Results

CT-aBMD was successfully measured in 976/978 hips (99.8%). A significant correlation was found between CT-aBMD and DXA-BMD ($r = 0.941$; $p < 0.001$). In the ROC analysis, the area under the curve to diagnose osteoporosis was 0.976. The diagnostic sensitivity and specificity were 88.9% and 96%, respectively, with the cutoff set at $0.625 \text{ g}/\text{cm}^2$.

Conclusion

Accurate DXA-BMD measurements and diagnosis of osteoporosis were performed from CT images using the system developed herein. As the models are open-source, clinicians can use the proposed system to screen osteoporosis and determine the surgical strategy for hip surgery.

Cite this article: *Bone Joint Res* 2023;12(9):590–597.

Keywords: Osteoporosis, Artificial intelligence, Bone mineral density

Article focus

- To develop a fully automated system to measure the bone mineral density (BMD) of the proximal femur from CT images.
- To validate the accuracy of the developed system in a multicentre study.

Key messages

- By using three deep learning-based models, the BMD of the proximal femur was accurately measured from CT images.
- The developed system can be used to opportunistically screen for osteoporosis.

Correspondence should be sent to Keisuke Uemura; email: surmountjp@gmail.com

doi: 10.1302/2046-3758.129.BJR-2023-0115.R1

Bone Joint Res 2023;12(9):590–597.

- The developed system can aid surgeons in selecting the type of implant for hip surgery.

Strengths and limitations

- A fully automated system to measure the BMD of the proximal femur from CT images was developed and validated in a large cohort.
- Open-source models were used for the analysis.
- More experiments may be needed to further confirm the robustness of the developed system.

Introduction

Screening and diagnosis of osteoporosis and subsequent adequate treatment are crucial for preventing fragile bone fractures that reduce quality of life. As recommended by the World Health Organization and several guidelines, osteoporosis is typically diagnosed by measuring the bone mineral density (BMD) of the proximal femur or lumbar region by dual-energy X-ray absorptiometry (DXA).¹⁻³ However, as osteoporosis is generally asymptomatic, patients are less likely to undergo DXA, leading to underdiagnosis and undertreatment of osteoporosis.⁴

To improve the diagnostic rate of osteoporosis and enable treatment before fragile bone fractures develop, several studies have examined the use of other imaging methods for screening osteoporosis.⁵ For example, screening via pulse-echo ultrasound of the lower leg,^{6,7} ultrasonography of the calcaneus (i.e. quantitative ultrasonography),⁸ and radiography of the metacarpal bone (i.e. microdensitometry) and wrist have been reported.^{8,9} Yet, the relatively low accuracy of these methods in diagnosing osteoporosis and limited availability of the equipment and methods necessary for clinical investigation remain an issue. Other screening methods include quantitative CT (QCT) images for BMD analysis. A commercially available software (e.g. QCT Pro; Mindways Software, USA) has been developed to estimate BMD measured by DXA (DXA-BMD) from CT images,¹⁰ and based on the very high correlation coefficient ($r > 0.9$) between areal BMD measured from QCT images and DXA-BMD,¹¹⁻¹⁵ CT-measured BMD is listed on the official positions by the International Society for Clinical Densitometry (ISCD).^{15,16} However, as commercial software adds to the cost, adoption of CT for measuring BMD and screening osteoporosis remains limited in clinical practice. Recently, an open-source software to measure the BMD from hip CT images was reported.¹⁷ However, the software was only verified in a small sample (75 cases) from one institution, and manual annotation of the landmarks at the proximal femur was necessary. Thus, developing an open-source fully automated system that can measure BMD from clinically acquired hip CT images was deemed necessary to help surgeons to select implants for hip arthroplasty (e.g. cemented or cementless stem) and suggest treatment for patients with osteoporosis. This study aimed to develop a fully automated system for quantifying proximal femur BMD from hip CT images and validate the

accuracy of the developed system in a multicentre study to confirm its versatility.

Methods

Participants. Ethical approval was obtained from the institutional review boards of all institutions participating in this retrospective study. The study participants consisted of a total of 857 patients (978 hips) who underwent hip surgery at three institutions (A: Osaka University; B: Yamaguchi University; and C: Ando Hospital) (Table I). The sample was composed of 189 male and 668 female patients, and the majority of the patients underwent surgery for osteoarthritis in institutions A and B, whereas it was for proximal femoral fracture in institution C. Patients who had metal implants in the spine or femur were excluded. Of the 978 hips, 377 (38.5%) were diagnosed as 'osteoporotic', according to the T-score for women calculated from the BMD measurement of the total proximal femoral region using DXA. These T-scores were standardized by a correction method proposed by the Japan Osteoporosis Society,¹ which refers to the guidelines of the ISCD.¹⁸

Image acquisition. Preoperative hip CT images acquired for surgical planning were analyzed in this study. During CT image acquisition, a calibration phantom made of urethane foam containing four hydroxyapatite rods (B-MAS200; Kyoto Kagaku, Japan) was placed under the patient's hip. For the three institutions, five CT models from three manufacturers were used for image acquisition (Table II). A standard clinical protocol with a matrix of 512×512 , a tube voltage of 120 kVP, and voxel sizes ranging from $(0.545 \text{ mm to } 0.977 \text{ mm}) \times (0.545 \text{ mm to } 0.977 \text{ mm}) \times (0.625 \text{ mm to } 2.0 \text{ mm})$ was used to acquire all CT images. The pixel size of the CT images was set to include the pelvis and femur within the region of interest (ROI), and the slice interval at institution C was discussed between each surgeon (KU, KT, HH) and the radiology department for their preference in the 3D reconstruction of the bone on a software with multiplanar planning (ZedHip; Lexi Co, Japan). A convolutional kernel representing bone or soft-tissue kernel was used to reconstruct CT images.

For each patient before or after the CT acquisition, DXA images of the proximal femur were acquired using three DXA models from two manufacturers (Table II). Specifically, for patients who underwent hip arthroplasty, DXA images were acquired preoperatively to determine the stem implant type (i.e. cement or cementless stem). Conversely, for patients with proximal femoral fracture, DXA was acquired postoperatively to determine the osteoporosis treatment method. As BMD values differ between the manufacturers, a previously reported equation was used to convert the value measured in GE's DXA (GE Healthcare Japan, Japan) to that measured in Hologic's DXA (Hologic Japan, Japan).¹⁹ The median interval between the acquisitions of CT and DXA images was eight days (interquartile range (IQR) 1 to 14).

Table I. Demographics of the patients from three institutions.

Demographic	A	B	C	Overall
Patients/hips, n (%)	315/315	167/288	375/375	857/978
Sex (male/female), n (%)	59/256	36/131	94/281	189/668
Mean age, yrs (SD)	57.3 (15.3)	65.5 (11.7)	82.6 (9.6)	70.0 (16.8)
Mean height, cm (SD)	157.9 (12.1)	154.9 (8.5)	153.5 (9.4)	155.4 (9.1)
Mean weight, kg (SD)	58.7 (12.7)	59.2 (12.7)	49.4 (10.4)	54.7 (12.5)
Mean BMI, kg/m ² (SD)	24.6 (4.3)	24.6 (4.3)	20.9 (3.7)	22.5 (4.2)
Hip disease, n of patients				
OA	263	148	2	413
ONFH	47	19		66
RDC	4			4
ALT	1			1
PFF			373	373
Osteoporosis, n of hips (%)	51 (16.2)	46 (16.0)	280 (74.7)	377 (38.5)

ALT, acetabular labrum tear; OA, osteoarthritis; ONFH, osteonecrosis of the femoral head; PFF, proximal femoral fracture; RDC, rapidly destructive coxopathy; SD, standard deviation.

Table II. CT images and dual-energy X-ray absorptiometry acquired at three institutions.

Factor	A	B	C
CT manufacturer (model)	GE (Optima CT660)	GE (Optima CT660)	Toshiba (Aquilion Precision)
CT convolution kernel	STD	STD	Siemens (SOMATOM Force)
CT voxel size, mm	(0.545 to 0.820) × (0.545 to 0.820) × (1.25)	(0.703 to 0.742) × (0.703 to 0.742) × (1.25)	Toshiba (Activion 16)
N of hips	315	6	Siemens (SOMATOM Force)
DXA manufacturer (model)	Hologic (Discovery A)	Hologic (Horizon A)	Toshiba (Activion 16)
Median CT–DXA duration, days (IQR)	7 (1 to 14)	0 (0 to 16)	GE (Revolution Maxima)

DXA, dual-energy X-ray absorptiometry; GE, General Electric Healthcare Japan; IQR, interquartile range; Toshiba, Toshiba Medical Systems.

Measurement methods. On the basis of a previous report, deep learning was used to develop a system to measure the BMD of the proximal femur.¹⁷ Specifically, models previously reported to segment the calibration phantom and the femur were used,^{17,20} and a new model to detect the landmarks of the proximal femur was developed and used (Figure 1).

Femur and phantom segmentation from CT images. A publicly available Bayesian U-Net model, which is a convolutional neural network for semantic segmentation that was previously reported and validated (Figure 1a), was used to perform femur and phantom segmentation from the CT images for each case.²¹ Briefly, the femoral segmentation model was trained on the manual segmentation of bilateral femora in 120 cases, and the phantom segmentation model was trained on the manual segmentation of the phantom in 40 cases. The accuracy of femoral segmentation was reported as a mean Dice coefficient²² of 0.985 (standard deviation (SD) 0.0065) and a mean symmetric surface distance (MSD)²³ of 0.175 mm (SD 0.084),²¹ whereas the accuracy of the phantom segmentation was reported as a median Dice coefficient of 0.977 (IQR 0.963 to 0.986) and a MSD of 0.116 mm (IQR 0.084–0.193).²⁰ Use of phantom segmentation enabled determination of a linear correlation equation between

the mean Hounsfield units (HUs) and the known densities of the phantom (0, 50, 100, 150, and 200 mg/cm³ in hydroxyapatite). Subsequently, using the correlation equation, the HUs of each voxel were converted into density (mg/cm³).

Development of a landmark detection model. To define the proximal femoral region for measuring BMD and account for the variations in patient positioning during CT, a deep-learning model that automatically selects the landmarks of the proximal femur from the CT images was developed. Specifically, to construct the training data, an orthopaedic surgeon with 15 years of experience (KU) manually selected nine landmarks: (1 to 2) 2 and 5 cm distal from the tip of the lesser trochanter; (3) head centre; (4) neck centre; (5 to 8) head-neck junction of the superior, anterior, posterior, and inferior regions; and (9) tip of the lesser trochanter, from CT images of institution A (315 hips) using software that allowed for multiplanar reconstruction (3D template; Kyocera, Japan). As pre-processing, CT images were cropped to include the proximal femoral region using the segmentation information of the femur and were downsampled to 128 × 128 × 128. During the training phase, data augmentation was conducted so that the model could account for the variations in patient positioning. Specifically, translation by (–25%,

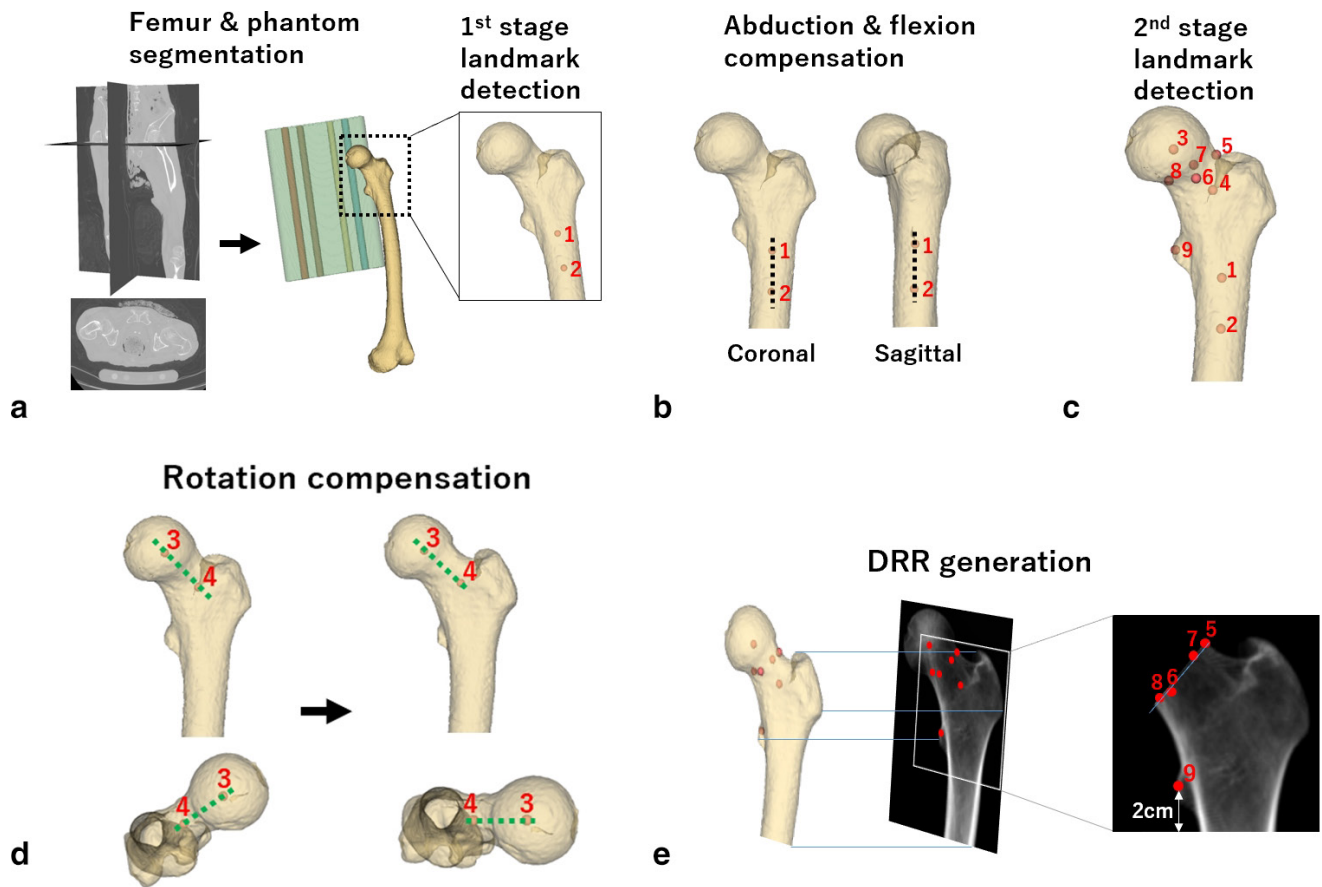


Fig. 1

Flowchart of the proposed method for measuring the CT-aBMD for a left femur. a) The femur and phantom were segmented from the CT images, and two landmarks ((1) 2 cm distal from the lesser trochanter; and (2) 5 cm distal from the lesser trochanter) were selected. b) The femoral volume was rotated to the neutral position on the coronal and sagittal planes. c) Additional seven landmarks were selected, namely: (3) head centre; (4) neck centre; (5 to 8) head-neck junction of the superior, anterior, posterior, and inferior regions; and (9) tip of the lesser trochanter. d) The femoral volume was rotated on the axial plane to the neutral position. e) The femoral volume in the neutral position was projected onto the coronal plane and was cropped using the projected landmarks to measure CT-aBMD. The black dotted line in b) indicates the shaft axis, and the green dotted line in d) indicates the neck axis. aBMD, areal bone mineral density; DRR, digitally reconstructed radiograph.

+25%) of the matrix size, rotation of (-30° , $+30^\circ$) around the x-axis, rotation of (-15° , $+15^\circ$) around the y- and z-axes, and scaling of (-20% , $+20\%$) were conducted.

CT-aBMD measurement. The landmarks were selected in two stages to account for substantial variability in patient positioning during CT. For example, patients from institution C had a large positioning variation because of pain caused by the fracture on the contralateral side. First, landmarks of 2 and 5 cm distal from the tip of the lesser trochanter were selected by the landmark detection model to account for femoral flexion and abduction during image acquisition (Figure 1a). Subsequently, the CT volume was rotated to the neutral position (Figure 1b), and landmark detection for the nine landmarks was performed again (Figure 1c). By using the landmarks of head and neck centres, the CT volume was rotated on the axial plane to account for rotation and compensate for femoral anteversion (Figure 1d). The rotated volume was projected onto the coronal plane using trilinear interpolation with a step length of 1 mm,²⁴ and a digitally

reconstructed radiograph (DRR) similar to the DXA images was generated (Figure 1e). The landmarks were also projected onto the coronal plane and used to isolate the proximal femoral region in the DRR (Figure 1e). The distal end of the proximal femur was defined as 2 cm distal from the tip of the lesser trochanter. The mean density of the proximal femoral region was measured and converted to units in DXA (g/cm^2) to enable comparison with DXA-BMD.¹⁷

Comparison of BMD measurements in CT and DXA. CT-aBMD correlated to DXA-BMD of the total proximal femur, and the correlation coefficients were calculated. Further, the ability of CT-aBMD to estimate the DXA-BMD was evaluated using the Bland–Altman plots, and absolute differences were also calculated. To diagnose osteoporosis (classified by the T-score in DXA-BMD), the receiver operating characteristic (ROC) curve was analyzed for CT-aBMD. As a subanalysis, the absolute differences between CT-aBMD and DXA-BMD were compared across institutions and CT manufacturers.

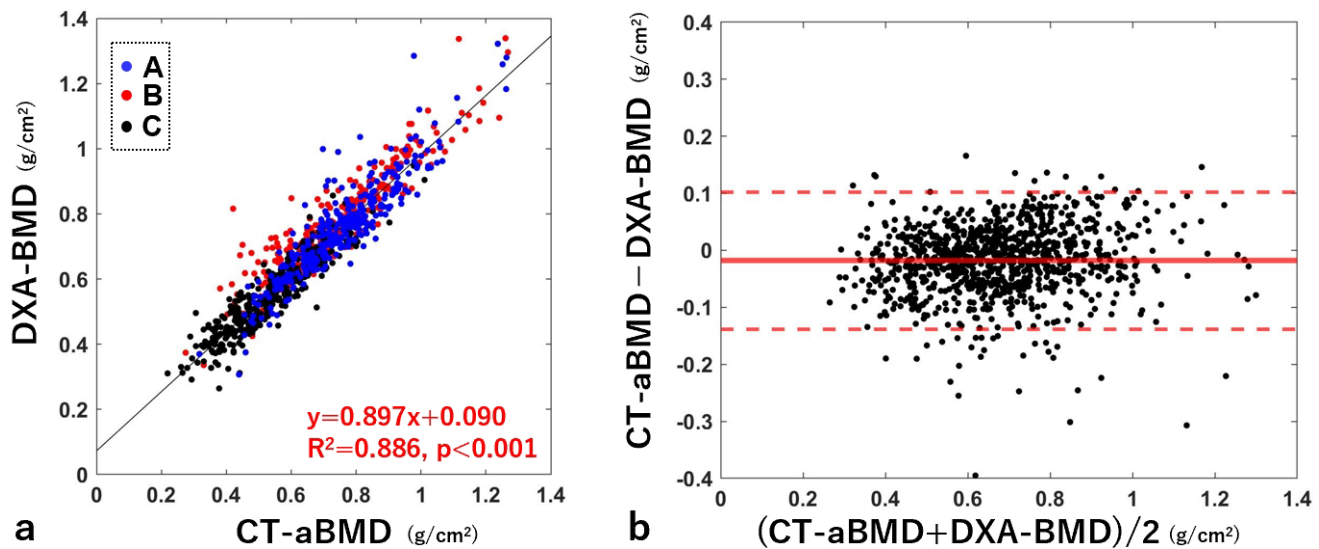


Fig. 2

a) Correlation plots between CT-aBMD and DXA-BMD and b) the corresponding Bland–Altman plot. For a), the blue, red, and black dots indicate cases from institutions A, B, and C, respectively. The black line indicates the regression line, and the red text indicates the regression equation, coefficient of determination, and p-value. The thick red line in b) indicates the mean value of the plots, and the thin red dotted lines indicate the 95% limits of agreement. aBMD, areal bone mineral density; DXA, dual-energy X-ray absorptiometry.

Accuracy of the landmark detection model and comparison of CT-aBMD measured from manually selected landmarks. The accuracy of the landmark detection model was validated internally and externally. For internal validation, fourfold cross-validation was performed on the training dataset (315 hips). For external validation, nine landmarks were manually selected for 103 hips acquired using Activion 16 (Toshiba Medical Systems, Japan) in institution C and reconstructed with a convolution kernel of FC30 (Table II).

By using the landmarks selected manually and automatically, the 3D distance between the landmarks was calculated. Further, the angles (abduction, flexion, and rotation) to rotate the images to the neutral position were calculated and compared. Finally, the CT-aBMD measured using the manually and automatically selected landmarks was compared.

Statistical analysis. The Shapiro–Wilk test was performed to confirm data normality. Normally distributed data were expressed as means (SDs), while non-normally distributed data were expressed as medians and IQRs. The correlation between two variables was assessed with a Pearson correlation coefficient, and $r > 0.8$ was considered to be very strong.²⁵ To diagnose osteoporosis, the area under the curve (AUC), optimal cutoff value, sensitivity, and specificity for the ROC curve analysis were calculated. The Kruskal–Wallis test was performed to compare absolute differences across institutions and CT manufacturers. All statistical analysis was performed using MATLAB v9.10 (MathWorks, USA), and p-values < 0.05 were considered to be statistically significant.

Results

CT-aBMD was successfully measured in 976/978 hips (99.8%). In one of the failed cases, CT-aBMD was not

measured because of an error in femoral segmentation; in the other case, CT-aBMD was not measured because of an error in landmark selection.

Comparison between DXA-BMD and CT-aBMD. A significant correlation was observed between CT-aBMD and DXA-BMD. The correlation coefficient was 0.941 ($p < 0.001$) (Figure 2a). In the Bland–Altman analysis, the mean difference between CT-aBMD and DXA-BMD was -0.02 g/cm^2 (Figure 2b). The median absolute error was 0.04 g/cm^2 (IQR 0.02 to 0.06).

Across institutions, the absolute error was significantly larger for institution B (0.05 g/cm^2) than for institutions A (0.03 g/cm^2) and C (0.03 g/cm^2) ($p = 0.002$ and $p < 0.001$, respectively (Kruskal–Wallis test); Supplementary Figure aa). Across CT manufacturers, the absolute error for GE scanner (0.03 g/cm^2) was significantly smaller than those for Toshiba (0.04 g/cm^2) and Siemens (0.05 g/cm^2) scanners ($p = 0.004$ and $p = 0.003$, respectively (Kruskal–Wallis test); Supplementary Figure ab).

ROC curve analysis. In the ROC analysis, the AUC to diagnose osteoporosis was 0.976 (Figure 3). The diagnostic sensitivity and specificity were 88.9% and 96%, respectively, with a cutoff value of 0.625 g/cm^2 .

Validation of the landmark detection model. The median 3D distance ranged from 2.0 mm to 7.9 mm for internal validation (Figure 4a). By contrast, the 3D distance ranged from 1.9 mm to 8.5 mm for external validation (Figure 4b). In the measurements of the hip angles from the landmarks, the median absolute angular difference calculated from manually and automatically selected landmarks was less than 2° for abduction and flexion for both internal and external validations. Moreover, the errors were 2.2° and 2.5° for the internal and external validations, respectively (Supplementary Figures ba and bb).

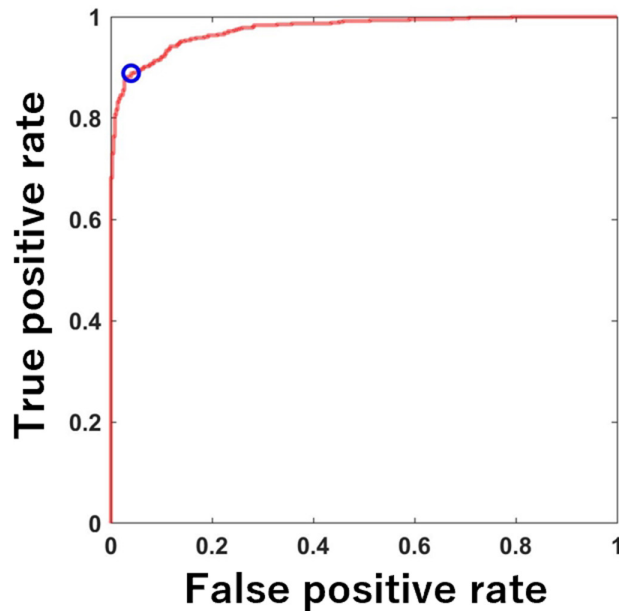


Fig. 3

Receiver operating characteristic curve analysis for diagnosing osteoporosis using CT-aBMD. The blue circle indicates the optimal cutoff point. aBMD, areal bone mineral density.

In the Bland–Altman analysis for CT-aBMD measurements using manually or automatically selected landmarks, the mean differences in CT-aBMD were -0.002 g/cm^2 and -0.007 g/cm^2 for the internal (Supplementary Figure bc) and external (Supplementary Figure bd) validations, respectively. The median absolute errors were 0.005 g/cm^2 (IQR 0.002 to 0.008) and 0.007 g/cm^2 (IQR 0.004 - 0.013) for the internal and external validations, respectively.

Discussion

In this study, a pipeline consisting of three deep-learning models was applied to develop an open-source system that automatically measures the BMD of the proximal femur from CT images.²⁶ A very strong correlation was observed between CT-aBMD and DXA-BMD ($r = 0.941$). In the ROC analysis for diagnosing osteoporosis, the AUC was 0.976 . Collectively, our results indicate that accurate DXA-BMD measurement and diagnosis of osteoporosis can be performed from clinical CT images using the system developed in this study.

Hips with measurement error. In the majority of hips (976 hips, 99.8%), CT-aBMD was automatically measured using the developed system. However, the system failed in quantifying the CT-aBMD in two cases. One case had an extremely low BMD (0.274 g/cm^2 , calculated as a T-score of -5.0)¹ that likely led to errors in femoral segmentation. In addition, when the landmarks were manually selected for these two cases as a post-hoc analysis, the hip angles relative to the CT table were 20.2° and 34.3° for flexion, 32.4° and 44.8° for adduction, and 61.7° and 73.8° for internal rotation. During CT, patient positioning was excessive and beyond the range of data augmentation

performed in the training dataset for segmenting the femur and detecting the landmarks; thus, errors likely occurred in these two cases.

Accuracy of the landmark detection model. The 3D distance errors in some landmarks were relatively large (e.g. 7.9 mm to 8.5 mm for 5 cm distal from the lesser trochanter; Figures 4a and 4b). Conversely, errors in measuring the CT-aBMD were small ($< 0.01 \text{ g/cm}^2$; Figures 4c and 4d). Previous studies have reported the importance of defining the distal border of the ROI in the BMD measurement, because a 1 mm change results in a BMD change of 0.54% to 0.68% .^{27,28} Further, the errors in angle measurements to account for the variance in patient positioning during CT were small (Figures 4c and 4d), especially for abduction and flexion, which independently lead to errors in BMD measurements.²⁹ As the distal border of the ROI was defined based on the tip of the lesser trochanter that had small errors in the landmark detection model (1.9 mm to 2.0 mm) and as hip abduction and flexion were accurately compensated, the CT-aBMD was likely quantified with less error.

Comparisons with previous reports. Previous studies have used commercially available software to measure the BMD of the proximal femur from CT images and reported very strong correlations with DXA-BMD, with correlation coefficients of 0.91 to 0.95 .^{11–15} In the ROC curve analysis, commercially available software can diagnose osteoporosis with sensitivity and specificity ranging from 83% to 99% .^{12,14,15}

In this study, the correlation coefficient between CT-aBMD and DXA-BMD was 0.941 , and the sensitivity and specificity for diagnosing osteoporosis were 88.9% and 96% , respectively. Interestingly, the cutoff value to diagnose osteoporosis was 0.625 g/cm^2 , strikingly similar to the value set for female Japanese patients using Hologic's DXA,¹ supporting the validity of our system in representing the DXA-BMD. Compared with the previous system requiring manual landmark selection, the correlation coefficient between CT-aBMD and DXA-BMD in the present study ($r = 0.941$) was comparable to that of another previous study analyzed for 75 cases ($r = 0.950$).¹⁷ Thus, the automated system developed herein had accuracy comparable to that of the previous method requiring manual input.

Overall, although areal BMD measured by our system and areal BMD measured by the commercially available software were not directly compared in this study, our results based on a large sample size from a multicentre study support the finding that the developed software has similar accuracy to the commercially available software in predicting DXA-BMD and screening for osteoporosis.

Comparison across institutions and CT manufacturers. Statistically significant differences were found in the absolute differences between CT-aBMD and DXA-BMD across institutions and CT manufacturers. Further investigation is necessary to clarify the causality behind the differences because direct comparison is not possible due to the variations in patient backgrounds across institutions and CT manufacturers. However, errors in predicting

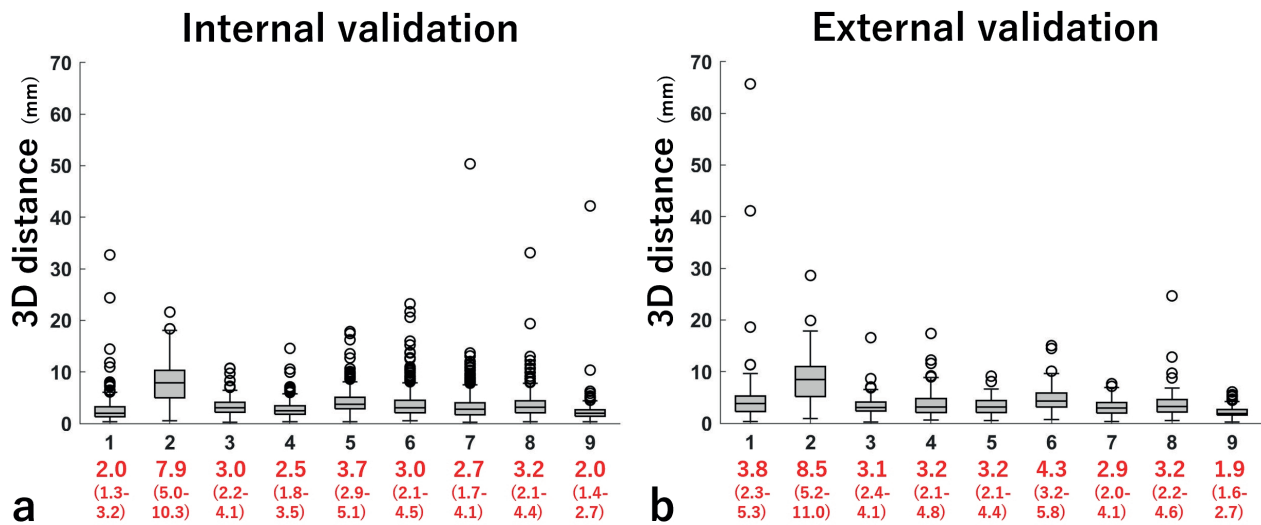


Fig. 4

Results of the 3D distance between the landmarks selected automatically and manually for a) internal and b) external validations. Red numbers indicate the median in millimetres (interquartile range) for landmarks 1 to 9 (landmark numbers refer to Figure 1c).

DXA-BMD using CT-aBMD were still small for the institution and manufacturer that had larger errors (0.05 g/cm^2) and had an excellent ability to detect osteoporosis.

Application of the developed system to clinical evaluation. The developed open-source system can be used to evaluate the proximal femur BMD and opportunistically screen for osteoporosis of patients who have acquired hip CT images for surgical planning of hip arthroplasty (e.g. CT-based navigation system and CT-based robotic surgery). The occurrence of periprosthetic fracture is an issue after cementless hip arthroplasty, so our system can help surgeons to select the stem implant type, such as cementless or cemented.³⁰


As a calibration phantom is not always included in CT images acquired for clinical investigation, additional experiment without phantom calibration (i.e. phantom-less) was performed as a post-hoc analysis. In the analysis, the correlation coefficient between CT-aBMD and DXA-BMD was 0.927 (Supplementary Figure ca), and the AUC for diagnosing osteoporosis was 0.967 (Supplementary Figure cb). The diagnostic sensitivity and specificity were 90.2% and 90.7%, respectively, with the cutoff value of 0.624 g/cm^2 . As the AUC, sensitivity, and specificity remained high, we believe that phantoms are not always necessary, and asynchronous calibration may be useful to screen CT images with notable differences in HUs.¹³ Further, we plan to apply the system to abdominal CT images acquired without a calibration phantom, and clarify the possibility of using the system for diagnosing osteoporosis from such images in a future study.

This study had some limitations. First, while our system was verified on a multicentre study using a large sample size (978 hips) in which the patient demographics, CT imaging, and reconstruction settings varied, errors may occur if CT images were acquired under a different

protocol and/or under different CT models. Applying our system and confirming its ability to accurately measure the BMD and screen for osteoporosis for CT images acquired under such conditions would be interesting. Second, patients with metal implants were excluded from the analysis because these artifacts may affect the results. As our next step, we intend to accurately measure the BMD of such patients and use a deep-learning method to reduce metal artifacts³¹ to provide adequate treatment of osteoporosis and prevent loosening of implants.

In conclusion, a fully automated system to predict the DXA-BMD from CT images was developed and verified in a multicentre study of 978 hips. A very strong correlation was found between CT-aBMD and DXA-BMD, with a correlation coefficient of 0.941. The median absolute error in estimating the DXA-BMD was 0.04 g/cm^2 . In the ROC analysis to diagnose osteoporosis, the AUC showed a high performance of 0.976. As the system can accurately measure BMD from CT images using open-source models,²⁶ clinicians can opportunistically screen osteoporosis from hip CT images and select the type of implant for hip surgery using the developed system.

Supplementary material

 Figures showing validation of the landmark detection model, comparison across the institutions and CT manufacturers, and results of the phantom-less bone mineral density measurements.

References

- Soen S, Fukunaga M, Sugimoto T, et al. Diagnostic criteria for primary osteoporosis: year 2012 revision. *J Bone Miner Metab.* 2013;31(3):247–257.
- Kanis JA, Cooper C, Rizzoli R, Reginster J-Y, Scientific Advisory Board of the European Society for Clinical and Economic Aspects of Osteoporosis (ESCEO) and the Committees of Scientific Advisors and National Societies of the International

- Osteoporosis Foundation (IOF).** European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int.* 2019;30(1):3–44.
3. **No authors listed.** Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: report of a WHO study group [Meeting held in Rome from 22 to 25 June 1992] World Health Organization. 1994. <https://apps.who.int/iris/handle/10665/39142> (date last accessed 11 August 2023).
 4. **Burden AM, Tanaka Y, Xu L, et al.** Osteoporosis case ascertainment strategies in European and Asian countries: a comparative review. *Osteoporos Int.* 2021;32(5):817–829.
 5. **Al-Hourani K, Tsang STJ, Simpson A.** Osteoporosis: current screening methods, novel techniques, and preoperative assessment of bone mineral density. *Bone Joint Res.* 2021;10(12):840–843.
 6. **Schousboe JT, Riekkinen O, Karjalainen J.** Prediction of hip osteoporosis by DXA using a novel pulse-echo ultrasound device. *Osteoporos Int.* 2017;28(1):85–93.
 7. **Dovjak P, Iglseider B, Rainer A, Dovjak G, Weber M, Pietschmann P.** Pulse-echo ultrasound measurement in osteoporosis screening: a pilot study in older patients. *Aging Clin Exp Res.* 2023;35(6):1221–1230.
 8. **Boonen S, Nijs J, Borghs H, Peeters H, Vanderschueren D, Luyten FP.** Identifying postmenopausal women with osteoporosis by calcaneal ultrasound, metacarpal digital X-ray radiogrammetry and phalangeal radiographic absorptiometry: a comparative study. *Osteoporos Int.* 2005;16(1):93–100.
 9. **Schmidutz F, Schopf C, Yan SG, Ahrend MD, Ihle C, Sprecher C.** Cortical bone thickness of the distal radius predicts the local bone mineral density. *Bone Joint Res.* 2021;10(12):820–829.
 10. **Engelke K.** Quantitative computed tomography - Current status and new developments. *J Clin Densitom.* 2017;20(3):309–321.
 11. **Khoo BCC, Brown K, Cann C, et al.** Comparison of QCT-derived and DXA-derived areal bone mineral density and T scores. *Osteoporos Int.* 2009;20(9):1539–1545.
 12. **Weber NK, Fidler JL, Keaveny TM, et al.** Validation of a CT-derived method for osteoporosis screening in IBD patients undergoing contrast-enhanced CT enterography. *Am J Gastroenterol.* 2014;109(3):401–408.
 13. **Pickhardt PJ, Bodeen G, Brett A, Brown JK, Binkley N.** Comparison of femoral neck BMD evaluation obtained using Lunar DXA and QCT with asynchronous calibration from CT colonography. *J Clin Densitom.* 2015;18(1):5–12.
 14. **Fidler JL, Murthy NS, Khosla S, et al.** Comprehensive assessment of osteoporosis and bone fragility with CT colonography. *Radiology.* 2016;278(1):172–180.
 15. **Ziemlewicz TJ, Maciejewski A, Binkley N, Brett AD, Brown JK, Pickhardt PJ.** Opportunistic quantitative CT bone mineral density measurement at the proximal femur using routine contrast-enhanced scans: Direct comparison with DXA in 355 adults. *J Bone Miner Res.* 2016;31(10):1835–1840.
 16. **Engelke K, Lang T, Khosla S, et al.** Clinical use of quantitative computed tomography (QCT) of the hip in the management of osteoporosis in adults: the 2015 ISCD Official Positions-Part I. *J Clin Densitom.* 2015;18(3):338–358.
 17. **Uemura K, Otake Y, Takao M, et al.** Development of an open-source measurement system to assess the areal bone mineral density of the proximal femur from clinical CT images. *Arch Osteoporos.* 2022;17(1):17.
 18. **Lewiecki EM, Binkley N, Morgan SL, et al.** Best practices for dual-energy X-ray absorptiometry measurement and reporting: International Society for Clinical Densitometry Guidance. *J Clin Densitom.* 2016;19(2):127–140.
 19. **Fan B, Lu Y, Genant H, Fuerst T, Shepherd J.** Does standardized BMD still remove differences between Hologic and GE-Lunar state-of-the-art DXA systems? *Osteoporos Int.* 2010;21(7):1227–1236.
 20. **Uemura K, Otake Y, Takao M, et al.** Automated segmentation of an intensity calibration phantom in clinical CT images using a convolutional neural network. *Int J Comput Assist Radiol Surg.* 2021;16(11):1855–1864.
 21. **Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N, Sato Y.** Automated muscle segmentation from clinical CT using Bayesian U-Net for personalized musculoskeletal modeling. *IEEE Trans Med Imaging.* 2020;39(4):1030–1040.
 22. **Dice LR.** Measures of the Amount of Ecologic Association Between Species. *Ecology.* 1945;26(3):297–302.
 23. **Styner M, Lee J, Chin B, et al.** The MIDAS Journal - MICCAI 2008 Workshop: MS Lesion Segmentation. 2008. <https://www.midasjournal.org/browse/publication/638> (date last accessed 9 August 2023).
 24. **Otake Y, Armand M, Armiger RS, et al.** Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: incorporation of fiducial-based C-arm tracking and GPU-acceleration. *IEEE Trans Med Imaging.* 2012;31(4):948–962.
 25. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.
 26. **Uemura K.** proximal_femur_BMD_measurement. GitHub. 2023. https://github.com/keisuke-uemura/proximal_femur_BMD_measurement (date last accessed 11 August 2023).
 27. **Feit A, Levin N, McNamara EA, et al.** Effect of positioning of the region of interest on bone density of the hip. *J Clin Densitom.* 2020;23(3):426–431.
 28. **Uemura K, Takao M, Otake Y, et al.** The effect of region of interest on measurement of bone mineral density of the proximal femur: Simulation analysis using CT images. *Calcif Tissue Int.* 2022;111(5):475–484.
 29. **Uemura K, Takao M, Otake Y, et al.** The effect of patient positioning on measurements of bone mineral density of the proximal femur: a simulation study using computed tomographic images. *Arch Osteoporos.* 2023;18(1):35.
 30. **Bukowski BR, Sandhu KP, Bernatz JT, et al.** CT required to perform robotic-assisted total hip arthroplasty can identify previously undiagnosed osteoporosis and guide femoral fixation strategy. *Bone Joint J.* 2023;105-B(3):254–260.
 31. **Sakamoto M, Hiasa Y, Otake Y, et al.** Bayesian segmentation of hip and thigh muscles in metal artifact-contaminated CT using convolutional neural network-enhanced normalized metal artifact reduction. *J Sign Process Syst.* 2020;92(3):335–344.

Author information:

- K. Uemura, MD, PhD, Associate Professor
- H. Hamada, MD, PhD, Associate Professor
- N. Sugano, MD, PhD, Professor
Department of Orthopaedic Medical Engineering, Graduate School of Medicine, Osaka University, Suita, Japan.
- Y. Otake, PhD, Associate Professor
- Y. Sato, PhD, Professor
Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan.
- K. Takashima, MD, PhD, Assistant Professor
- S. Okada, MD, PhD, Professor
Department of Orthopaedics, Graduate School of Medicine, Osaka University, Suita, Japan.
- T. Imagama, MD, PhD, Associate Professor
- T. Sakai, MD, PhD, Professor
Department of Orthopaedics, Graduate School of Medicine, Yamaguchi University, Ube, Japan.
- M. Takao, MD, PhD, Professor, Department of Bone and Joint Surgery, Graduate School of Medicine, Ehime University, Toon, Japan.

Author contributions:

- K. Uemura: Conceptualization, Methodology, Project administration, Software, Formal analysis, Data curation, Writing – original draft, Visualization, Funding acquisition.
- Y. Otake: Resources, Methodology, Writing – review & editing, Funding acquisition.
- K. Takashima: Writing – review & editing.
- H. Hamada: Writing – review & editing.
- T. Imagama: Data curation, Resources, Writing – review & editing, Supervision.
- M. Takao: Resources, Writing – review & editing, Supervision.
- T. Sakai: Resources, Writing – review & editing, Supervision.
- Y. Sato: Resources, Supervision, Funding acquisition.
- S. Okada: Resources, Supervision.
- N. Sugano: Resources, Writing – review & editing, Supervision.

Funding statement:

- The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: funding from the 2022 Japan Osteoporosis Foundation Grant for Bone Research Sponsored by Amgen K.K., Japanese Orthopaedic Association Research Grant (JOA-Subsidized Science Project Research: 2023-2), and Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) numbers 19H01176, 20H04550, and 21K16655.

ICMJE COI statement:

- The authors declare that they have no conflicts of interest.

Data sharing:

- The datasets generated and analyzed in the current study are not publicly available due to data protection regulations. Access to data is limited to the researchers who have obtained permission for data processing. Further inquiries can be made to the corresponding author.

Acknowledgements:

- We thank Dr Mazen Soufi, Dr Sotaro Kono, and Mr Tomoki Hakotani for their help in the analysis. We also thank Dr Tatsuya Kitaura for his help in the acquisition of data. We thank Enago (www.enago.jp) for their English language editing services.

Ethical review statement:

- Approval for this study was obtained from the Institutional Review Board at each institution, and informed consent was obtained from all patients involved in the study in the form of opt-out.

Open access funding:

- The authors report that the open access funding for their manuscript was self-funded.

© 2023 Author(s) et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original author and source are credited.