Estimating Bone Mineral Density and Muscle Mass from EOS Low Dose X-ray Imaging System

Kazuki Suehara[1], Yi Gu[1], Yoshito Otake[1], Keisuke Uemura[2], Masashi Okamoto[3,4], Kunihiko Tokunaga[4], Hugues Talbot[5], and Yoshinobu Sato[1]

[1] Nara Institute of Science and Technology
[2] Osaka University
[3] Kameda Daiichi Hospital
[4] Niigata University
[5] CentraleSupélec, Université Paris-Saclay, France
{gu.yi.gu4,otake,yoshi}@is.naist.jp

Abstract. The EOS imaging system is a low-dose, biplanar X-ray modality offering high-fidelity anatomical visualization in standing and seated positions, benefiting total hip arthroplasty (THA) by providing accurate skeletal alignment and implant positioning pre- and postoperatively. Evaluating bone mineral density (BMD) and muscle mass before surgery is useful for predicting outcomes and tailoring rehabilitation. Although CT and DXA can assess these metrics effectively, they increase cost and radiation exposure. Recent advances in deep learning have enabled BMD and muscle mass estimation from plain radiographs, among which one promising approach with potentially high generalizability to new modality utilized 2D–3D registration with CT of the same patient in training data preparation. However, limited EOS availability constrains large data collection. We devised and validated a deep learning framework to predict BMD and muscle mass from EOS images by fine-tuning a model trained on plain radiographs. Our dataset comprised 77 pairs of pre- and postoperative EOS images and CT scans, then underwent 2D–3D registration to create paired training data. Our contribution is two-fold: 1) we achieved reliable BMD and muscle mass estimation in THA cases with minimal training data, and 2) we experimentally demonstrated that only 40 paired EOS–CT images were sufficient to reach high accuracy, supporting feasibility in resource-limited settings. Future work will extend this approach to broader patient populations and anatomical sites while performing external validation to assess potential domain shifts across different facilities.

Keywords: EOS Imaging · Bone Mineral Density · Muscle Mass

1 Introduction

Bones and muscles are essential to human mobility, and conditions such as osteoporosis and sarcopenia threaten quality of life among older adults [3, 10]. Osteoporosis, marked by low bone mass, raises fracture risk and healthcare costs,

while sarcopenia worsens frailty. Early detection is vital for timely interventions. As shown in Figure 1, the EOS imaging system employs a moving X-ray source that travels vertically behind the patient, enabling biplanar, full-body imaging at low radiation doses [2]. This differs from conventional X-ray imaging, which typically uses a fixed source emitting radiation radially. By simultaneously capturing frontal and lateral views, EOS facilitates three-dimensional (3D) reconstructions of anatomical structures and more comprehensive analyses of spinal and lower-limb alignment.

Recent research has highlighted the utility of conventional X-ray images in opportunistic screening for osteoporosis, fueled by advances in deep learning. Plain radiographs are widely accessible, cost-effective, and entail lower radiation compared with computed tomography, making them suitable for large-scale bone mineral density (BMD) assessments. Convolutional neural networks, for instance, can automatically extract cortical and trabecular bone features from simple X-rays, showing promising accuracy in predicting BMD [7, 8, 1]. Among them, one promising approach with potentially high generalizability to new modality utilized 2D–3D registration with CT of the same patient in training data preparation [4]. These studies suggest that plain radiographs may offer a practical alternative to dual-energy X-ray absorptiometry (DXA), currently the clinical gold standard for osteoporosis diagnosis.

In this work, we propose a novel pipeline that focuses on robust BMD and muscle mass estimation in total hip arthroplasty (THA) cases using minimal training data. Our contribution is two-fold: (1) we achieved reliable BMD and muscle mass predictions with only a small dataset by fine-tuning the pre-trained model using the method proposed in [4], and (2) we experimentally demonstrated that merely 40 paired EOS–CT images suffice to obtain high accuracy. This finding underscores the feasibility of our approach in resource-limited environments, where acquiring large imaging datasets is often challenging. By leveraging the low-dose capability of EOS imaging and integrating advanced machine learning techniques, our method holds promise for more frequent and safer patient monitoring, ultimately supporting earlier diagnoses and improved outcomes for THA patients at risk of osteoporosis or sarcopenia.

2 Methods

2.1 Dataset

In this study, we collected datasets from two institutions (Table 1). The first dataset was gathered from 600 patients (484 females, 116 males) who underwent total hip arthroplasty (THA) at Institution 1. For each patient, we collected multiple X-ray images of the hip (in abduction, adduction, standing, supine, and resting positions) as well as lower-limb CT scans. The patient's height was $157.5 \pm 8.3cm$, weight $58.7 \pm 12.2kg$, and BMI $23.6 \pm 4.0kg/m^2$. This dataset provides a broad spectrum of pre- and postoperative images, offering valuable anatomical diversity for training and evaluating algorithms aimed at BMD and muscle mass estimation.

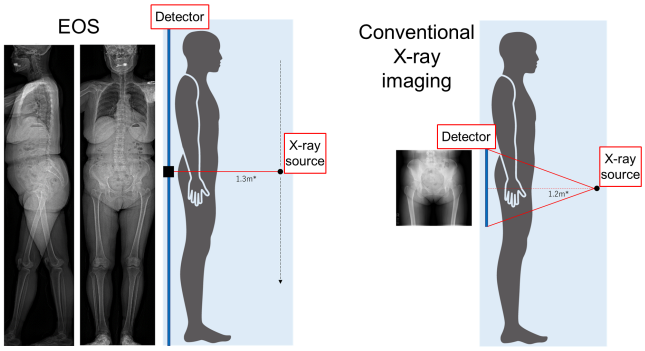

Fig. 1. Differences between the EOS imaging system and conventional X-ray imaging. Conventional X-ray imaging emits radiation radially from a fixed X-ray source, whereas the EOS imaging system uses a moving source to emit radiation horizontally, enabling full-body imaging.

The second dataset originated from Institution 2, comprising 77 patients (70 females, 7 males) who were scheduled to receive THA and had no other implants in either hip prior to surgery. In each case, we collected full-body EOS images in the standing position and complementary lower-limb CT scans. This group had a height of $153.8 \pm 8.0cm$, weight of $59.2 \pm 9.9kg$, and BMI of $25.0 \pm 3.7kg/m^2$. Because EOS imaging delivers low-dose, biplanar radiographs, it permits frequent postoperative monitoring with minimal radiation exposure.
Using the two datasets, we followed the DRR generation steps described in prior work [4] to create standardized projections of the proximal femur. These synthetic radiographs were then used to develop and validate our proposed framework for BMD and muscle mass evaluation under varying clinical conditions.

Table 1. Details of the dataset used in this study

Acquired institution	Institution 1		Institution 2	
Purpose	pre-training		fine-tuning and testing	
Modality	CT	X-ray	CT	EOS
Data dimension	3D	2D	3D	2D
# of patients	600	600	77	77
# of images	600	2461[a]	77	77
Field of view	Lower extremity	Hip	Pevis to knee	Whole body

[a] 4-5 X-ray images per patient

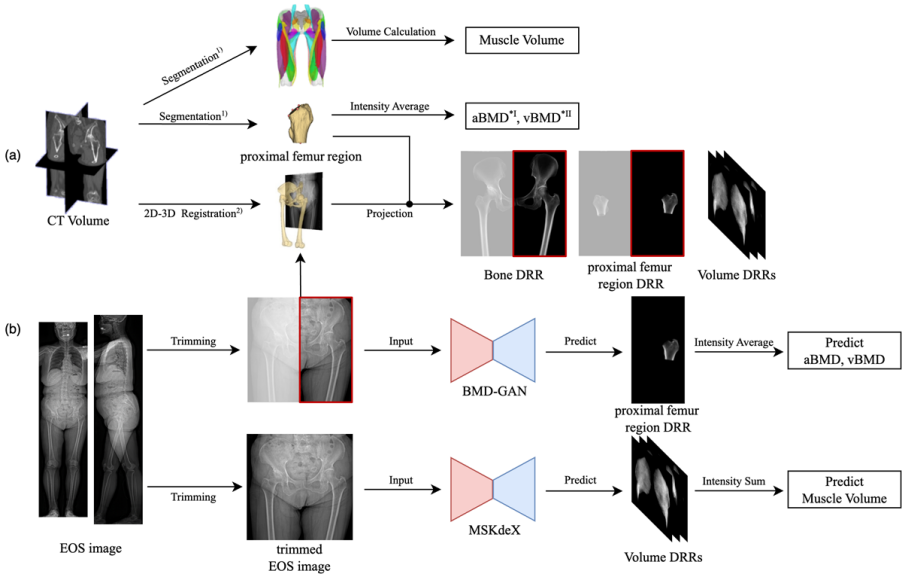

Fig. 2. Overview of Bone Density and Muscle Volume Estimation:(a) Ground Truth from CT Images:Bone density (aBMD, vBMD) and muscle volume are obtained from CT images using segmentation, 2D-3D registration, and DRR generation. Muscle volume is calculated from segmented regions, and bone density is averaged from the proximal femur.(b) Estimation from EOS Images:Trimmed anterior EOS images are input into deep learning models—BMD-GAN for bone density and MSKdeX for muscle volume. Predicted DRRs are processed to estimate bone density and muscle volume.

2.2 QCT Calibration Protocol

Accurate BMD reference values were obtained from quantitative CT (QCT) scans by following two different calibration strategies for each institution, which depend on whether an external calibration phantom was present.

(i) Phantom-based calibration for the pre-training cohort. All 600 CT volumes collected at Institution 1 were acquired with the B-MAS200 phantom (Kyoto Kagaku Inc., Kyoto, Japan) placed beneath the patient's hip. Voxel intensities were converted to the hydroxyapatite equivalent density using a linear calibration derived from the intensity of the phantom inserts. The calibration pipeline was validated in a prior study, yielding r = 0.95 and SEE = 0.050 g/cm^{-2} for aBMD over the proximal femur [11].

(ii) Phantom-less calibration for the EOS fine-tuning cohort. The 77 CT scans from Institution 2 were acquired without a calibration phantom but with the same scanner model, tube potential (120 kVp), and reconstruction kernel as the phantom scans. We therefore adopted a fixed scanner-specific calibration curve obtained by averaging the per-scan slopes and intercepts measured in our prior study. This "phantom-less" approach has been shown, under protocol-

matched conditions, to retain high agreement with DXA ($r > 0.94$ for aBMD of the femoral neck in 86 subjects [11]).

2.3 Proposed Framework and Pre-training

In this study, we constructed a framework for bone density and muscle mass estimation by adapting methods from previous research [4, 5], originally developed for X-ray images, to EOS data. As illustrated in Figure 2(a), we first utilized a trained U-net model [6] to obtain musculoskeletal segmentation labels and femoral landmark points from CT images, which serve as ground-truth references for bone mineral density and muscle mass calculations. Next, we perform 2D–3D registration [9] to align the CT and EOS images, enabling the creation of digitally reconstructed radiographs (DRRs) aligned with EOS images. During DRR generation, we utilize CT scans from the pelvis to the knee and extract the proximal femoral region based on landmark detection—replicating the imaging range typically used in DXA examinations.
For bone density estimation, we treat each patient's left and right hip regions as separate data samples, effectively splitting the dataset into two images per patient. In contrast, muscle mass estimation leverages the patient's entire dataset as a single input as shown in Figure 2(b). By training and predicting with the two models (predicting BMD and muscle mass) on these EOS images, we establish a system capable of estimating both BMD and muscle mass.

2.4 Experimental Setup and Evaluation Metrics

We employed a 5-fold cross-validation strategy to assess the performance of our proposed framework. In each fold, the training set comprised 61 or 62 patients, while the test set included the remaining 15 or 16 patients. For BMD estimation, we split each patient's data into two images (left and right hip), effectively doubling the number of training samples; by contrast, for muscle mass estimation, we used the entire image of each patient as a single input. We used only frontal images from EOS in this experiment. We compared four experimental setups using the same training settings (e.g., loss function and optimizer) for pre-training and fine-tuning:

- **Train**: A model trained from scratch on EOS data.
- **Pre-trained**: A model initially trained on conventional X-ray images without further fine-tuning.
- **Linear calibration**: The same X-ray-based model whose predictions are corrected using a simple linear regression approach.
- **Fine-tuning**: The X-ray-based model, further refined using EOS images.

We evaluated each method against CT-derived ground truths using three metrics—Pearson Correlation Coefficient (PCC), Intraclass Correlation Coefficient (ICC), and Mean Absolute Error (MAE)—to gauge accuracy in BMD and muscle mass predictions.

3 Results

3.1 Bone Mineral Density Estimation

Table 2 shows the results of the experiment. The experiments revealed that fine-tuning a model originally trained on conventional X-ray images to EOS data yields the highest accuracy for bone mineral density (BMD) estimation. Overall, the model trained from scratch on EOS images (Train) achieved the lowest performance, likely due to the small training set—approximately 120 images—which limited its capacity to learn robust representations. In contrast, the model pre-trained solely on X-ray images (Pre-trained) produced a Pearson correlation coefficient (PCC) of 0.874, indicating a moderate linear relationship. However, its intraclass correlation coefficient (ICC) was notably poor (-0.604), suggesting low consistency in individual-level predictions. This discrepancy may stem from brightness mismatches between the original X-ray data and EOS images.
When the Pre-trained model was refined further with EOS images (Fine-tuning), both the PCC (0.916) and ICC (0.894) increased substantially. These findings underscore the importance of domain-specific adaptation, as the fine-tuned model consistently outperformed the other approaches across all evaluation metrics.

Table 2. BMD and Muscle Mass Estimation Results (best and second best results were marked by bold and underline, respectively)

Model type	BMD estimation	Muscle mass estimation				
		Glu. max	Glu. med.	Glu. min.	Iliacus	Obt. ext.
PCC						
Train	0.791	0.705	0.753	0.63	0.792	0.701
Pre-trained	<u>0.874</u>	**0.821**	<u>0.839</u>	<u>0.692</u>	<u>0.823</u>	<u>0.766</u>
Linear calibration	0.869	<u>0.798</u>	0.832	0.670	0.805	0.740
Fine-tuning	**0.916**	0.795	**0.840**	**0.746**	**0.843**	**0.798**
ICC						
Train	0.710	0.483	0.655	0.596	0.721	0.647
Pre-trained	-0.604	0.344	0.568	0.579	-0.218	0.698
Linear calibration	<u>0.861</u>	**0.784**	**0.821**	<u>0.629</u>	**0.795**	<u>0.719</u>
Fine-tuning	**0.894**	<u>0.631</u>	<u>0.772</u>	**0.729**	<u>0.763</u>	**0.762**

3.2 Muscle Mass Estimation

In muscle mass estimation, the model trained from scratch on EOS data (Train) again showed the lowest accuracy, likely due to the relatively small training set. Table 2 compares prediction performance across different muscles, indicating that both the Pre-trained and Fine-tuning models surpass the Train model in overall consistency. Notably, Fine-tuning outperforms Pre-trained by exhibiting higher ICC for the majority of targeted muscles. This suggests that tailoring a

model initially trained on conventional X-ray images to EOS data substantially enhances predictive reliability.
Among the examined muscles, the iliopsoas muscle demonstrated the largest gains with Fine-tuning. The improvements in accuracy highlight the value of domain-specific adaptation when transitioning from plain X-ray images to EOS-based datasets. By leveraging the strengths of an existing pre-trained model and refining its parameters on new, low-dose image data, Fine-tuning emerges as the most viable option for robust muscle mass estimation. These findings reinforce the conclusion that, while starting from scratch in a small-scale dataset can be suboptimal, a well-curated transfer learning approach yields more accurate and consistent clinical measurements.

3.3 Comparison between pre- and post-operative images

Figure 3 illustrates the comparison between pre- and postoperative measurements in terms of bone mineral density (BMD). In Figure 3(a), the horizontal axis represents the preoperative BMD values calculated from CT scans, while the vertical axis corresponds to their postoperative counterparts. A Pearson correlation coefficient of 0.995 indicates an extremely high agreement, suggesting that our ground truth values are reliable.
To further assess reproducibility, we calculated the coefficient of variation (CV) for each patient, under the assumption that BMD remains unchanged during the short interval between the pre- and postoperative exams. As shown in Figure 3(b), the root mean square CV (RMS-CV) for the CT-based ground truth was 1.01%, confirming the reliability of the ground truth. The Pre-trained model's reproducibility was bad (about 10.01%), while it improves quite a lot by Fine-tuning (2.67%).
Figure 3(c) compares absolute errors between the predicted and ground-truth BMD values in both the preoperative and postoperative scenarios, displayed as boxplots. These results indicate that the magnitude of prediction error is similar in both cases, suggesting that our method's performance remains consistent regardless of surgical intervention. Taken together, these findings demonstrate that the proposed framework maintains high accuracy and reliability in monitoring BMD before and after THA.

3.4 Effect of Training Data Size

We further investigated how varying the quantity of training data for fine-tuning affects BMD inference accuracy. As illustrated in Figure 4, we began with data from 10 patients—equivalent to 20 total images—randomly selected for fine-tuning. We then successively added data from 10 more patients in each iteration, observing how incremental increases influenced predictive performance. The results indicate that once 40 images (i.e., data from 20 patients) are included, inference accuracy stabilizes. Beyond roughly 80 images, both the intraclass correlation coefficient (ICC) and mean absolute error (MAE) remain at a high,

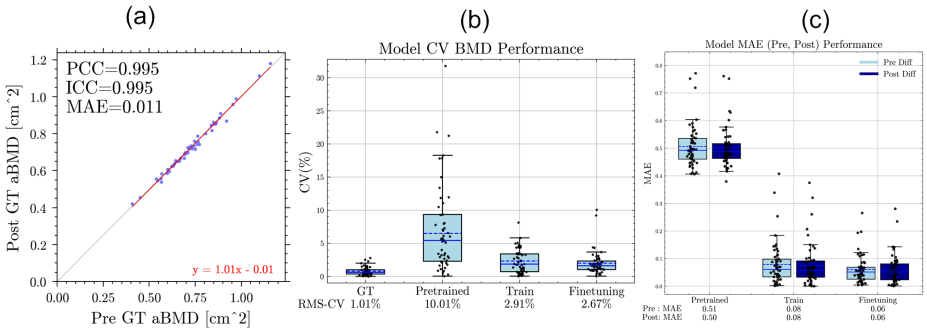

Fig. 3. BMD estimated from pre- and post-operative CT and EOS images. (a) scatter plot of aBMD, showing high correlation (PCC = 0.995, ICC = 0.995) between measurements from preoperative (x-axis) and postoperative (y-axis) CT images, confirming accurate measurement. (b) covariance of variation between pre- and post-operative CT and EOS, (c) mean absolute error (MAE) between ground truth (CT) and predicted (EOS) aBMD for pre- and post-operative stages. Box plots show consistent prediction accuracy for Train and Fine-tuning, indicating reliable performance, while Pre-trained showed a much larger error.

consistent level, suggesting diminishing returns from adding additional training samples.

4 Discussion

This study demonstrated that the proposed pipeline can reliably estimate BMD and muscle mass using limited amounts of EOS imaging data, especially through the application of transfer learning. Fine-tuning models pre-trained on conventional X-ray images consistently outperformed those trained from scratch, underscoring the value of leveraging existing knowledge to compensate for the small size of EOS datasets. Notably, our experiments showed that fewer than 40 paired EOS–CT images can suffice to achieve stable performance and that increasing the training data beyond 80 images yields diminishing returns. Additionally, we observed minimal differences in estimation accuracy between pre- and postoperative scenarios, suggesting that the method remains robust across short-term changes in patient anatomy.

Importantly, our approach is not intended to supplant DXA in population-level osteoporosis screening; rather, it capitalises on the use of data that are already collected during routine orthopaedic care. High-volume arthroplasty centres acquire standing EOS scans for alignment analysis, implant templating, and postoperative monitoring; mining these low-dose images for BMD and muscle metrics therefore adds zero radiation, no extra scan time, and no incremental cost. This opportunistic workflow is particularly valuable for frail or mobility-limited patients who struggle with the repositioning required for DXA or conventional

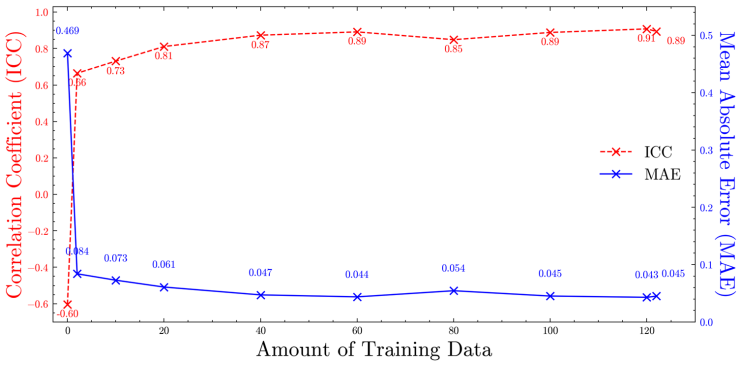

Fig. 4. Impact of training data size on prediction accuracy, showing the effect of varying the amount of training data on prediction accuracy during fine-tuning. Data was incrementally added by selecting 10 patients (20 images) at a time. Additionally, results averaged from training with one patient at a time (10 iterations) and five patients at a time (2 iterations) are included. The results indicate that prediction accuracy stabilizes with 40 images, and similar ICC and MAE values are achieved with 80 images or more.

CT, and it enables longitudinal musculoskeletal assessment without additional imaging burden. By turning a pre-existing image into a multifaceted biomarker source, the proposed method complements DXA rather than competes with it, extending bone-health surveillance into clinical encounters where it was previously absent.

While our results are promising, several limitations merit attention. First, the dataset primarily consisted of patients from one or two institutions, potentially restricting the generalizability of our model to other clinical settings or demographics. Second, BMD and muscle mass calculations relied on CT scans as the reference standard, which may not be universally available or practical for all patients. Third, the small overall sample size, despite our cross-validation approach, could introduce biases in model training and evaluation. Finally, variations in patient positioning, EOS image quality, and acquisition protocols present additional challenges when generalizing the results across diverse healthcare environments.

Moving forward, expanding the dataset to include multi-institutional and multi-ethnic populations would strengthen the robustness and applicability of the proposed framework. Incorporating advanced data augmentation techniques, domain adaptation strategies, or synthetic image generation might further enhance model performance, especially in low-resource settings. Exploring time-series analyses—tracking BMD and muscle mass changes longitudinally—could offer a more dynamic and clinically relevant perspective on patient outcomes.

Acknowledgments. This study was funded by MEXT/JSPS KAKENHI (19H01176, 20H04550, 21K16655, 23K15714) and AMED under Grant Number JP25hma322015.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anam, M., Hussain, M., Nadeem, M.W., Javed Awan, M., Goh, H.G., Qadeer, S., et al.: Osteoporosis prediction for trabecular bone using machine learning: a review. Computers, Materials & Continua (CMC) **67**(1) (2021)
2. Brage, K., Mussmann, B., Geijer, M., Larsen, P., Jensen, J.: Clinical application of eos imaging system: a scoping review protocol. JBI evidence synthesis **21**(5), 1009–1015 (2023)
3. Cruz-Jentoft, A.J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A.A., et al.: Sarcopenia: revised european consensus on definition and diagnosis. Age and ageing **48**(1), 16–31 (2019)
4. Gu, Y., Otake, Y., Uemura, K., Soufi, M., Takao, M., Talbot, H., Okada, S., Sugano, N., Sato, Y.: Bone mineral density estimation from a plain x-ray image by learning decomposition into projections of bone-segmented computed tomography. Medical Image Analysis **90**, 102970 (2023)
5. Gu, Y., Otake, Y., Uemura, K., Takao, M., Soufi, M., Hiasa, Y., Talbot, H., Okada, S., Sugano, N., Sato, Y.: Mskdex: musculoskeletal (msk) decomposition from an x-ray image for fine-grained estimation of lean muscle mass and muscle volume. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 497–507. Springer (2023)
6. 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 transactions on medical imaging **39**(4), 1030–1040 (2019)
7. Hsieh, C.I., Zheng, K., Lin, C., Mei, L., Lu, L., Li, W., Chen, F.P., Wang, Y., Zhou, X., Wang, F., et al.: Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. Nature communications **12**(1), 5472 (2021)
8. Kim, S., Kim, B.R., Chae, H.D., Lee, J., Ye, S.J., Kim, D.H., Hong, S.H., Choi, J.Y., Yoo, H.J.: Deep radiomics–based approach to the diagnosis of osteoporosis using hip radiographs. Radiology: Artificial Intelligence **4**(4), e210212 (2022)
9. Otake, Y., Armand, M., Armiger, R.S., Kutzer, M.D., Basafa, E., Kazanzides, P., Taylor, R.H.: Intraoperative image-based multiview 2d/3d registration for image-guided orthopaedic surgery: incorporation of fiducial-based c-arm tracking and gpu-acceleration. IEEE transactions on medical imaging **31**(4), 948–962 (2011)
10. Sander, B., Elliot-Gibson, V., Beaton, D.E., Bogoch, E.R., Maetzel, A.: A coordinator program in post-fracture osteoporosis management improves outcomes and saves costs. JBJS **90**(6), 1197–1205 (2008)
11. Uemura, K., Otake, Y., Takao, M., Makino, H., Soufi, M., Iwasa, M., Sugano, N., Sato, Y.: Development of an open-source measurement system to assess the areal bone mineral density of the proximal femur from clinical ct images. Archives of Osteoporosis **17**(1), 17 (2022)