

# Automated osteoporosis classification and *T*-score prediction using hip radiographs via deep learning algorithm

Yu-Pin Chen\* , Wing P. Chan\*, Han-Wei Zhang\*, Zhi-Ren Tsai\*, Hsiao-Ching Peng, Shu-Wei Huang, Yeu-Chai Jang and Yi-Jie Kuo 

## Abstract

**Background:** Despite being the gold standard for diagnosing osteoporosis, dual-energy X-ray absorptiometry (DXA) is an underutilized screening tool for osteoporosis.

**Objectives:** This study proposed and validated a controllable feature layer of a convolutional neural network (CNN) model with a preprocessing image algorithm to classify osteoporosis and predict *T*-score on the proximal hip region via simple hip radiographs.

**Design:** This was a single-center, retrospective study.

**Methods:** An image dataset of 3460 unilateral hip images from 1730 patients (age  $\geq 50$  years) was retrospectively collected with matched DXA assessment for *T*-score for the targeted proximal hip regions to train (2473 unilateral hip images from 1430 patients) and test (497 unilateral hip images from 300 patients) the proposed CNN model. All images were processed with a fully automated CNN model, X1AI-Osteo.

**Results:** The proposed screening tool illustrated a better performance (sensitivity: 97.2%; specificity: 95.6%; positive predictive value: 95.7%; negative predictive value: 97.1%; area under the curve: 0.96) than the open-sourced CNN models in predicting osteoporosis. Moreover, when combining variables, including age, body mass index, and sex as features in the training metric, there was high consistency in the *T*-score on the targeted hip regions between the proposed CNN model and the DXA ( $r=0.996$ ,  $p < 0.001$ ).

**Conclusion:** The proposed CNN model may identify osteoporosis and predict *T*-scores on the targeted hip regions from simple hip radiographs with high accuracy, highlighting the future application for population-based opportunistic osteoporosis screening with low cost and high adaptability for a broader population at risk.

**Trial registration:** TMU-JIRB N201909036.

**Keywords:** deep learning, neural network, osteoporosis, radiographs, *T*-score

Received: 4 August 2023; revised manuscript accepted: 9 February 2024.

## Introduction

Osteoporosis is a common systemic skeletal disorder that leads to low bone mass and increased risk of fragility fractures.<sup>1</sup> Hip fracture is the most debilitating among all fragility fractures, resulting in chronic pain, loss of independence,<sup>2</sup> decreased quality of life,<sup>3</sup> and high mortality following hip fracture surgery.<sup>4</sup> The classification of osteoporosis is defined by the lowest bone mineral density (BMD) on the axial bone, including spine and

bilateral hip regions.<sup>5</sup> However, BMD on the proximal hip area is especially critical to directly reflect the future risk of hip fracture.<sup>6</sup> Dual-energy X-ray absorptiometry (DXA) is the golden standard and most extensively used method for BMD measurement in the hip and spine regions.<sup>5</sup> Nevertheless, to screen osteoporosis, DXA lacks the minimum service requirement, as the geographic availability and associated utilization of DXA are inadequate, especially for rural

*Ther Adv Musculoskelet Dis*

2024, Vol. 16: 1–11

DOI: 10.1177/  
1759720X241237872

© The Author(s), 2024.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:  
**Yi-Jie Kuo**  
Department of  
Orthopedics, Wan Fang  
Hospital, Taipei Medical  
University, No. 111, Sec. 3,  
Xinglong Road, Wenshan,  
Taipei 11696, Taiwan  
(R.O.C.)

Department of  
Orthopedics, School of  
Medicine, College of  
Medicine, Taipei Medical  
University, Taipei, Taiwan  
[benkuo5@tmu.edu.tw](mailto:benkuo5@tmu.edu.tw)

**Yu-Pin Chen**  
Department of  
Orthopedics, Wan Fang  
Hospital, Taipei Medical  
University, Taipei City,  
Taiwan

Department of  
Orthopedics, School of  
Medicine, College of  
Medicine, Taipei Medical  
University, Taipei City,  
Taiwan

**Wing P. Chan**  
Department of Radiology,  
Wan Fang Hospital, Taipei  
Medical University, Taipei  
City, Taiwan

Department of Radiology,  
School of Medicine,  
College of Medicine, Taipei  
Medical University, Taipei  
City, Taiwan

**Han-Wei Zhang**  
Biomedica Corporation,  
New Taipei City, Taiwan  
Program for Aging,  
China Medical University,  
Taichung City, Taiwan

Institute of Population  
Health Sciences, National  
Health Research  
Institutes, Miaoli County,  
Taiwan

Department of Electrical and Computer Engineering, Institute of Electrical Control Engineering, National Yang Ming Chiao Tung University, Hsinchu City, Hsinchu County, Taiwan

**Zhi-Ren Tsai**

Department of Computer Science and Information Engineering, Asia University, Taichung City, Taiwan

Department of Medical Research, China Medical University Hospital, China Medical University, Taichung City, Taiwan

Center for Precision Medicine Research, Asia University, Taichung City, Taiwan

**Hsiao-Ching Peng**

Biomedica Corporation, New Taipei City, Taiwan

**Shu-Wei Huang**

Department of Applied Science, National Taitung University, Taitung City, Taitung County, Taiwan

**Yeu-Chai Jang**

Department of Obstetrics and Gynecology, Wan Fang Hospital, Taipei Medical University, Taipei City, Taiwan

\*These authors contributed equally

residents.<sup>7</sup> It is of utmost significance to develop other reliable and easily accessible methods to identify the risk of osteoporosis at the hip besides DXA assessment.

The hip radiographs may be informative for screening osteoporosis. The Singh index (SI), a six-graded classification system for bone density of the proximal femoral neck based on the visibility of the trabecular types and arrangement, is a simple grading system for diagnosing osteoporosis with plain radiographs.<sup>8</sup> The SI directly reflects the osteoporotic grading of the proximal femur and potentially predicts the future risk of hip fracture; however, it is highly subjective by clinicians with fair inter- and intra-observer agreement.<sup>9,10</sup> Recently, deep learning (DL) algorithms have illustrated remarkable progress in developing a screening tool for osteoporosis based on simple hip radiographs<sup>11–15</sup>; however, these were limited to using DL for only identifying osteoporosis rather than further predicting the *T*-score on the targeted region owing to the inappropriate object detection results with methodological flaws.

Hence, this study aims to propose and validate a fully automated convolutional neural network (CNN) model (X1AI-Osteo) to (i) segment the bony contour of the proximal hip as the region of interests (ROIs), (ii) classify osteoporosis, and (iii) directly predict the *T*-score on the proximal femur from a single hip radiograph (Figure 1). Furthermore, the classification of osteoporosis and predicted *T*-score by DL are compared with the results assessed by DXA on the targeted hip to validate the clinical reliability and applicability of this automated osteoporosis screening tool.

## Methods

### Study design

This single-center, retrospective study investigated the diagnostic accuracy of osteoporosis using the proposed X1AI-Osteo model and compared it with two CNN models (InceptionV3, ResNet50). The Ethics Committee of Taipei Medical University approved this study (registration number: TMU-JIRB N201909036). Owing to the retrospective nature of the study and the analysis of anonymous clinical data, the Ethics Committee waived the need for obtaining informed consent.

### Patient selection

The data with both simple hip anterior–posterior (AP) radiographs and DXA examination results of consecutive patients who presented to one medical center between November 2017 and September 2019 were retrospectively reviewed. The inclusion criteria were as follows: (i) aged  $\geq 50$  years and (ii) underwent both hip AP radiography and DXA within 6 months. We excluded hip radiographs on the left or right side of hips containing image-analyzing obstacles, including retained metal implants and severe osteoarthritis and osteonecrosis of the femoral head or foreign body materials, from DL model training. The reporting of this study conforms to the Strengthening the Reporting of Observational Studies in Epidemiology statement<sup>16</sup> (Supplemental Material).

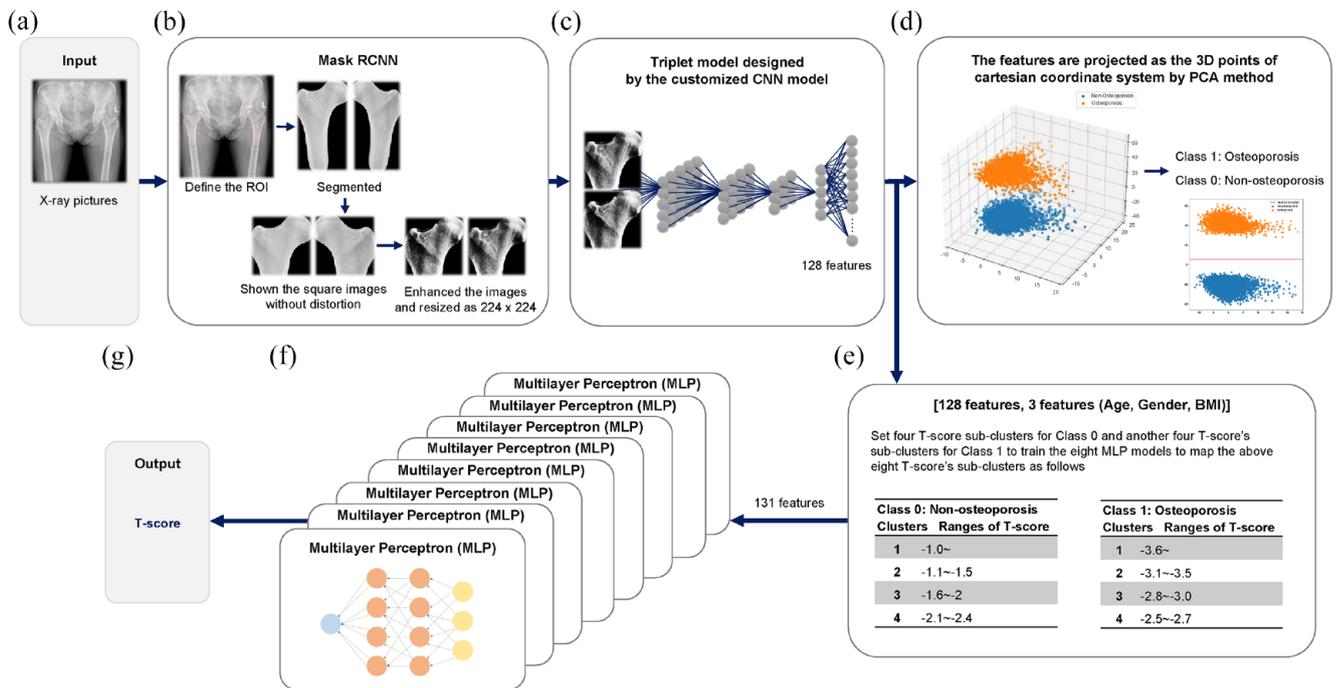
The lowest *T*-score on the targeted proximal hips, including total hip and femoral neck regions for bilateral sides of the hip, from each patient assessed by DXA were obtained to classify osteoporosis on the targeted hip and train the proposed CNN model. We defined the diagnosis of osteoporosis and osteopenia as *T*-score  $\leq -2.5$  and  $-1.5$ , respectively.<sup>1</sup> The *T*-score measurement by DXA was done using the Lunar Prodigy Advance System (GE Healthcare, WI, USA).

### Anatomical segmentation

Figure 1(a) defines the analyzing area on the hip radiographs and Figure 1(b) defines the ROIs for the bony contours of the proximal femur. A pair of ROI images were predicted and segmented by the Mask RCNN model [Figure 1(b)]. Next, the square images were used to predict osteoporosis and *T*-score, which were resized as  $224 \times 224$ , also defined in Figure 1(b), without distortion, and were cropped from the ROI images. To highlight the texture of the femoral neck, the enhanced images were obtained from the square images with/without mirror process by contrast enhancement techniques, which were the training/validating/testing data for the inputs of the triplet model.

### Training

To make triplet model learning more targeted, we proposed customized CNN model, shown in Figure 1(c) and Supplemental Figure S1, to obtain the critical 128 features of each sub-image



**Figure 1.** Schematic representation of the workflow for osteoporosis and bone mineral density estimation using a radiograph. (a) Definition of the analyzing area on the hip radiographs. (b) Definition of the ROIs for the bony contours of the proximal femur. A pair of ROI images were predicted and segmented by the Mask RCNN model. (c) The triplet model designed by the customized CNN model and obtain the critical 128 features of each sub-image based on the labeled target. (d) Two-dimensional projection features by principal component analysis dimensionality reduction for 131 triplet features to obtain two clusters of features separated by the triplet algorithm. Two clusters of points for two classes (Classes 0–1) were classified by the red classification boundary. (e) Eight *T*-score's sub-clusters were selected by the given reasonable ranges. (f) Training eight MLP (multilayer perceptron neural network) models and mapping the abovementioned eight *T*-score sub-clusters using genetic algorithm. (g) Calculate the *T*-score. CNN, convolutional neural network; MLP, multilayer perceptron neural network; RCNN, region-based convolutional neural network; ROI, region of interest.

based on the labeled target: Class 1, osteoporosis target; Class 0, non-osteoporosis target. Furthermore, we use principal component analysis dimensionality reduction for 131 triplet features to obtain these two clusters of features separated by the triplet algorithm; Figure 1(d) and Supplemental Figure S2 show the clustering result. Moreover, these two clusters of points for two classes (Classes 0–1) were classified by the red classification boundary [Figure 1(d) and Supplemental Figure S2]. Then, the eight *T*-score's sub-clusters were selected by the given reasonable ranges [Supplemental Table S1 and Figure 1(e)].

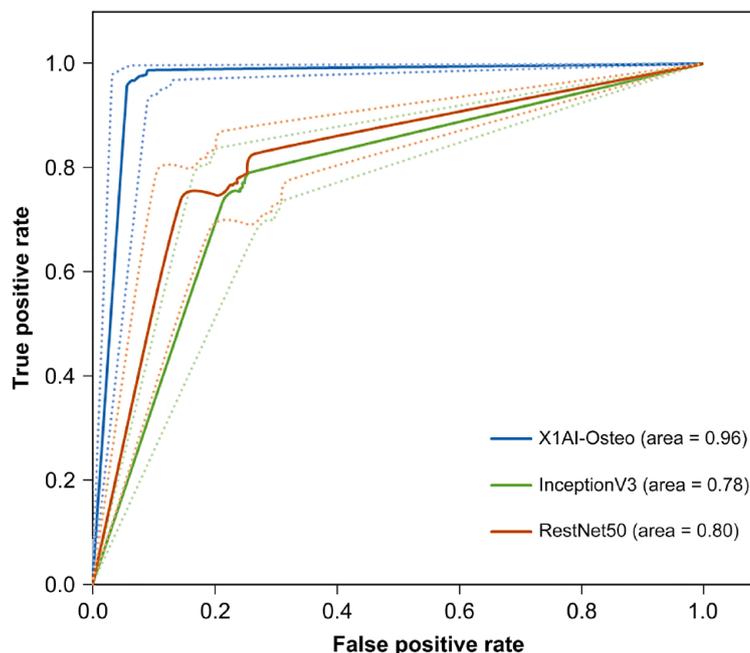
Supplemental Table S1 shows the two cases we designed. Case 1: If the Class 0's testing data were closest to one of the four classes mapped to one of the four *T*-score sub-clusters of Class 0, then the inference of this testing data was the output of that class's MLP model, and the fifth class was the outliers for 131 features of Class 0, which were filtered by multi-linear regression. Case 2: If

the Class 1's testing data were closest to one of the four classes mapped to one of the four *T*-score sub-clusters of Class 1, then the inference of this testing data was the output of that class's MLP model, and the fifth class was the outliers for 131 features of Class 1, which were filtered by multi-linear regression, and shown in Supplemental Figure S3 and Supplemental Table S1.

Next, the regression errors (*reg\_err*) of all 2D points (Supplemental Figure S2) were calculated by the regression output (*reg\_out*) compared with the desired *T*-score value (*tsv*), as shown in Supplemental Figure S3. This study defined the *reg\_err* of *tsv* as:

$$\text{reg\_err} = \text{tsv} - \text{reg\_out}$$

The outliers are subject to the filter condition  $|\text{reg\_err}| \geq 0.5$ . After filtering the outliers, the clusters 1–4 data were reserved (Supplemental Figure S3).



**Figure 2.** Receiver–operating characteristic curves for each osteoporosis prediction tool. All performance measures were evaluated on the test dataset, imputed separately.

To calculate the precise  $T$ -score shown in Figure 1(g), we set four  $T$ -score sub-clusters for Class 0 and another four  $T$ -score sub-clusters for Class 1 (Supplemental Table S1) to train the eight MLP models [Figure 1(f)] and map the abovementioned eight  $T$ -score sub-clusters using genetic algorithm (GA).<sup>17</sup> If the testing data were closest to one of eight sub-clusters, then the inference of this testing data was the output ( $T$ -score) of one of the MLP models. In technology, we selected the learning rate; neurons of the first, second, and third layers; and the training epochs as the searched hyper-parameters for GA to optimize the three layers of eight MLP models. Supplemental Figure S4 shows the learning curves. After the eight MLP models for eight groups were trained (Supplemental Table S2), we decided on the 131 features input of Class 0 or Class 1 in MLP models to evaluate the  $T$ -score value.

#### Model evaluation

All performance measures were evaluated only on the test dataset, imputed separately. The receiver–operating characteristic (ROC) curve in Supplemental Figure S5 is based on the different data augmentation implemented using the black and white background images, the different

degree of angles near the vertical bone direction, and different hyper and model parameters for the proposed CNN structure. Figure 2 summarizes the ROC curves based on the different CNN structures for triplet models.

The guided Grad-CAM (Gradient-weighted Class Activation Mapping) provided a direct visualization of the values in a map and combined the Grad-CAM and back-propagation visualization techniques. It showed information significant for classification – the high gradient of the input to the last convolutional layer. In this study, the heatmap visualizations were displayed relative to the range of values in the image. All visualizations were performed using iridescent map projections. Within the ROI, high attenuation was shown in green and low attenuation in red; hence, this study selected the best model.

## Results

#### Data source

Overall, 1730 patients [age: >50 years; mean age: 72.4 (standard deviation (SD) 11.1) years; 1332 (77.0%) female] with concomitant hip radiograph and DXA examination within 6 months were enrolled. As a simple hip radiograph contained

bilateral sides of hip images for analysis, 3460 unilateral hip images were collected for training and testing of X1AI-Osteo. We excluded 490 unilateral hip images (249 right hips and 241 left hips) owing to image-analyzing obstacles, including retained metal implants or bony deformity. Finally, 2473 unilateral hip images from 1430 patients were utilized for the training set and 497 unilateral hip images from 300 patients for the testing set.

Table 1 presents the subjects' characteristics. The mean *T*-score assessed by DXA was  $-2.6$  (SD 1.1). The DXA identified 1045 (60.4%) patients with osteoporosis based on the lowest *T*-score on the targeted hip region. The mean interval between DXA and a hip radiograph was 38.8 (SD 73.2) days.

#### *Prediction performance for osteoporosis*

Currently, the contour segmentation model will mark out three types: (1) screws, (2) joints, and (3) femur. Our program only extracts the contour of the femur, while screws and joints (image-analyzing obstacles) are not processed. Before calculating the *T*-score, medical personnel confirm the absence of distortion in segmented images.

The performance of Segmentation in terms of Intersection over Union (IoU) is as follows (Table 2). The contour detection model uses 2506 images, including 163 bone nail images, 137 artificial joint images, and 2206 femoral images. Standards for IoU values: Artificial joints and femur must be greater than 0.97, the style of bone nails must be diverse and complicated, and the IoU value must be greater than 0.95.

The default features of pre-trained CNNs are not suitable for osteoporosis. Therefore, the triplet method is needed to fine-tune the default features of pretrained models. The fine-tuned features can highlight the more obvious differences between the characteristics of osteoporosis and non-osteoporosis, which is more helpful for the convergence of the proposed classifier (triplet model). The segmented images need to be resized to inputs of resolution size  $224 \times 224 \times 3$  for the proposed model, the ResNet50, and InceptionV3 models.<sup>18,19</sup>

Table 3 presents the performance assessment of the three CNN models (ResNet50, InceptionV3, and X1AI-Osteo) applying the 497 test sets of unilateral hip radiographs. Among the

three prediction tools, X1AI-Osteo exhibited the superior performance in predicting osteoporosis [sensitivity: 97.2%; specificity: 95.6%; positive predictive value (PPV): 95.7%; negative predictive value (NPV): 97.1%], followed by ResNet50 (sensitivity: 84.3%; specificity: 75.8%; PPV: 77.8%; NPV: 82.8%) and InspectionV3 (sensitivity: 78.3%; specificity: 78.6%; PPV: 78.6%; NPV: 78.3%). Figure 2 demonstrates the higher area under the curve (AUC) in predicting hip osteoporosis using X1AI-Osteo than that by ResNet50 and InspectionV3 (AUC: 96.4%, 80.1%, and 78.5%, respectively).

#### *Prediction performance for T-score*

Figure 3 summarizes the X1AI-Osteo performance to predict BMD. The Pearson's correlation coefficient between DXA-measured and X1AI-Osteo-predicted *T*-score was 0.996 ( $p < 0.001$ ). Supplemental Figure S6 demonstrates high concordance correlation coefficient [0.996 (95% confidence interval (CI): 0.995–0.997)] between DXA-measured and X1AI-Osteo-predicted *T*-score. Figure 4 confirms the high consistency of *T*-score prediction using X1AI-Osteo than that by DXA measurement.

#### *Visualization of the ROI by X1AI-Osteo*

Figure 5 presents the focused visualization area attained by guided Grad-CAM. In the radiographs of patients with and without osteoporosis, the relatively distinct areas of shading (obscure trabeculation) on the femoral neck and trochanteric region were identified as deep-learned feature areas.

## **Discussion**

Mostly, osteoporosis progresses silently until fragility fractures happen, highlighting the clinical importance of osteoporosis screening programs.<sup>20</sup> Considering the limited availability of DXA as a screening modality for osteoporosis,<sup>7</sup> DL-based modalities using a simple hip radiograph in opportunistic osteoporosis screening constitute a potential domain despite technical and clinical concerns.<sup>11,12,14,15</sup> In our model, the performance on osteoporosis classification was robust with DXA as a reference and competent with (or even superior to) the performance of the existing DL-based osteoporosis screening tools using simple hip radiographs.<sup>11,12,14,15</sup> The proposed and customized CNN network (X1AI-Osteo) rather

**Table 1.** Characteristics of the study population by XIAI-Osteo prediction tool input variables and DXA-based bone imaging biomarkers.

Input variable	Overall <sup>a</sup> , n (%)			Right hip neck T-score <sup>b</sup> , n (%)			Left hip neck T-score <sup>b</sup> , n (%)			Overall neck T-score <sup>b</sup> , n (%)			
	Normal	Osteopenia	Osteoporosis	Normal	Osteopenia	Osteoporosis	Excluding <sup>c</sup>	Normal	Osteopenia	Osteoporosis	Excluding <sup>c</sup>	Normal	Osteopenia
Overall <sup>a</sup>	1730 (100)	223 (12.9)	492 (28.4)	766 (44.3)	249 (14.4)	202 (11.7)	241 (13.9)	512 (29.6)	775 (44.8)	241 (13.9)	175 (10.1)	510 (29.5)	1045 (60.4)
T-score, mean (SD)	-2.6 (1.1)	-0.8 (0.6)	-2.0 (0.3)	-3.2 (0.5)	-	-0.8 (0.6)	-	-2.0 (0.3)	-3.2 (0.9)	-	-0.8 (0.6)	-2.0 (0.3)	-3.3 (0.8)
DXA input variables													
Age group (years), mean (SD)	72.4 (11.1)	66.5 (10.0)	67.9 (10.0)	74.7 (10.6)	79.6 (9.7)	66.6 (10.1)	79.0 (10.5)	68.3 (10.0)	74.7 (10.6)	79.0 (10.5)	66.2 (9.8)	68.3 (10.3)	75.5 (10.7)
50-59	231 (13.4)	62 (27.8)	109 (22.2)	55 (7.2)	5 (2.0)	59 (29.2)	10 (4.1)	107 (20.9)	55 (7.1)	10 (4.1)	51 (29.1)	111 (21.8)	69 (6.6)
60-69	514 (29.7)	83 (37.2)	188 (38.2)	209 (27.3)	34 (13.7)	71 (35.1)	41 (17.0)	187 (36.5)	215 (27.7)	41 (17.0)	65 (37.1)	184 (36.1)	265 (25.4)
70-79	474 (27.4)	51 (22.9)	128 (26.0)	222 (29.0)	73 (29.3)	47 (23.3)	62 (25.7)	144 (28.1)	221 (28.5)	62 (25.7)	41 (23.4)	135 (26.5)	298 (28.5)
≥80	511 (29.5)	27 (12.1)	67 (13.6)	280 (36.6)	137 (55.0)	25 (12.4)	128 (53.1)	74 (14.5)	284 (36.6)	128 (53.1)	18 (10.3)	80 (15.7)	413 (39.5)
Sex													
Men	398 (23.0)	68 (30.5)	88 (17.9)	171 (22.3)	71 (28.5)	63 (31.2)	69 (28.6)	103 (20.1)	163 (21.0)	69 (28.6)	61 (34.9)	99 (19.4)	238 (22.8)
Women	1332 (77.0)	155 (69.5)	404 (82.1)	595 (77.7)	178 (71.5)	139 (68.8)	172 (71.4)	409 (79.9)	612 (79.0)	172 (71.4)	114 (65.1)	411 (80.6)	807 (77.2)
Height, mean (SD)	155.6 (8.1)	159.6 (8.5)	156.4 (7.3)	153.7 (7.9)	155.9 (8.6)	159.2 (8.7)	155.9 (9.1)	156.8 (7.5)	153.7 (7.6)	155.9 (9.1)	160.0 (8.5)	157.2 (7.6)	154.0 (7.9)
Weight, mean (SD)	57.4 (10.8)	64.1 (11.2)	59.8 (10.3)	54.4 (9.5)	55.9 (11.4)	65.0 (11.9)	56.2 (11.1)	59.7 (9.9)	54.2 (9.6)	56.2 (11.1)	65.0 (11.5)	60.4 (10.5)	54.6 (9.7)
BMI <sup>d</sup>													
Obese	304 (17.6)	56 (25.1)	119 (24.2)	101 (13.2)	28 (11.2)	62 (30.7)	38 (15.8)	108 (21.1)	96 (12.4)	38 (15.8)	48 (27.4)	119 (23.3)	137 (13.1)
Overweight	411 (23.8)	76 (34.1)	116 (23.6)	161 (21.0)	58 (23.3)	65 (32.2)	48 (19.9)	130 (25.4)	168 (21.7)	48 (19.9)	58 (33.1)	132 (25.9)	221 (21.1)
Normal	901 (52.1)	90 (40.4)	240 (48.8)	435 (56.8)	136 (54.6)	74 (36.6)	132 (54.8)	258 (50.4)	437 (56.4)	132 (54.8)	68 (38.9)	244 (47.8)	589 (56.4)
Underweight	114 (6.6)	1 (0.4)	17 (3.5)	69 (9.0)	27 (10.8)	1 (0.5)	23 (9.5)	16 (3.1)	74 (9.5)	23 (9.5)	1 (0.6)	15 (2.9)	98 (9.4)
Interval between DXA and hip radiograph, mean (SD)	38.8 (73.2)	45.7 (86.6)	33.5 (68.4)	40.4 (72.5)	38.2 (70.8)	39.4 (80.9)	42.5 (69.3)	35.0 (71.4)	40.1 (73.4)	42.5 (69.3)	42.7 (85.2)	34.1 (69.1)	40.5 (72.8)

<sup>a</sup>The entire study population (training + test datasets).

<sup>b</sup>The T-score values were as follows: normal (>-1.5), osteopenia (<=-1.5 and >-2.5), osteoporosis (<=-2.5).

<sup>c</sup>Excluded owing to image-analyzing obstacles, including retained metal implants or bony deformity.

<sup>d</sup>The BMI values were as follows: obese (≥27); overweight (≥24 and <27); normal (≥18.5 and <24); underweight (<18.5). BMI, body mass index; DXA, dual-energy X-ray absorptiometry; SD, standard deviation.

**Table 2.** The performance of segmentation in terms of IoU.

Segmentation	Image	Standards IoU	Average IoU	True	True percentage (%)	False	False percentage (%)
Bone nail	163	0.95	0.968	160	98.2	3	1.8
Artificial joint	137	0.97	0.987	135	98.5	2	1.5
Femur	2206	0.97	0.991	2202	99.8	4	0.2

IoU, intersection over union.

**Table 3.** Discriminatory performance (%) of the osteoporosis AI prediction tools.

Discriminatory measures <sup>a</sup>	InceptionV3	ResNet50	X1AI-Osteo
AUC (95% CI)	78.5 (74.3–82.6)	80.1 (76.0–84.1)	96.4 (94.5–98.3)
Sensitivity (95% CI)	78.3 (72.9–83.1)	84.3 (79.5–88.5)	97.2 (94.6–98.8)
Specificity (95% CI)	78.6 (73.2–83.4)	75.8 (70.2–80.9)	95.6 (92.5–97.7)
PPV (95% CI)	78.6 (73.2–83.4)	77.8 (72.6–82.5)	95.7 (92.7–97.7)
NPV (95% CI)	78.3 (72.9–83.1)	82.8 (77.6–87.3)	97.1 (94.5–98.8)

Analysis is based on the test set, which comprised 300 individuals (497 unilateral hip radiographs). The CIs were calculated using the bootstraps, as detailed in section 'Methods'.

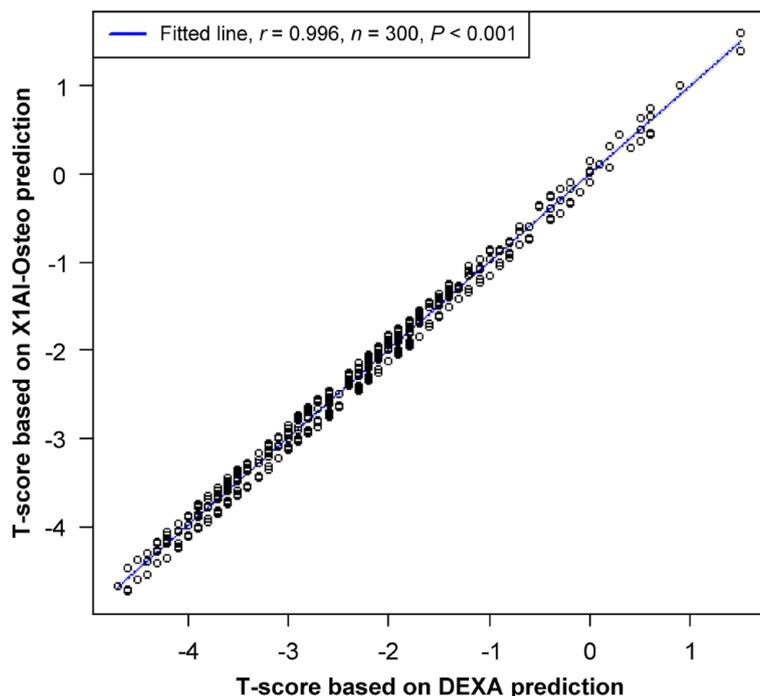
<sup>a</sup>All measures were evaluated and averaged across the 11 imputed datasets of the test sets.

AI, artificial intelligence; AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

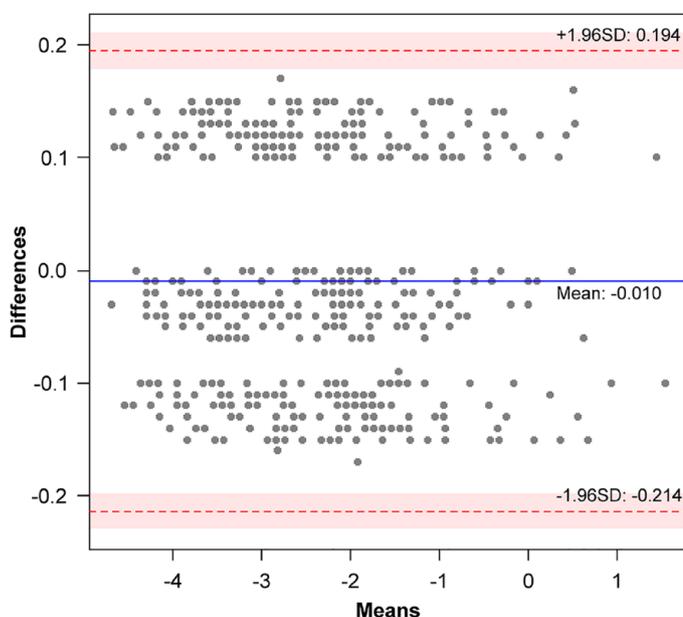
than the open-source CNN models utilized in our screening tool and the introduction of the automated segmentation of bony contour, as well as the image-enhancing process on the proximal femur, primarily attributed to its excellent performance in the diagnosis of osteoporosis. Our study proposes the aligned data augmentation of small data based on the data need reduction by the image process method to manage the overfitting issue of small data to train the proposed CNN model. Compared with the two most popular CNN models (InceptionV3 and ResNet50), our model illustrated a more compact model architecture (with faster model convergence to speed up about 12 times), resulting in better interpretation performance in predicting osteoporosis. Besides, our study introduced the method of automated segmentation of the bony contour on the proximal femur and in the workflow of the DL model, which can not only avoid the potential bias due to manual annotation of ROI but also simplify the analysis process in the clinical application. Our study introduced an effective method of automated segmentation of the bony contour on the

proximal femur and in the workflow of the DL model, building upon methodologies such as those explored in the paper 'Deep Radiomics-based Approach to the Diagnosis of Osteoporosis Using Hip Radiographs.' We acknowledge the contributions of previous studies in this field. Of note, the automated segmentation of the bony contour can facilitate focusing on the visualization area of the DL feature on the proximal femur cortex and trabecular patterns of the neck and trochanter region (Figure 5), with the main analyzing area for BMD by DXA examination<sup>21</sup> and also matched the rationales of SI in diagnosing osteoporosis based on plain hip radiographs.<sup>8</sup> Therefore, our model may create the added values of the published DL models<sup>11,12</sup> based on the image process method for enhancing the texture of hip radiographs, which would be extremely valuable in a clinical setting.

To the best of our knowledge, this is the first study to compare the single model in predicting *T*-score with the multi-cluster models. In addition, the positive results obtained rationalize



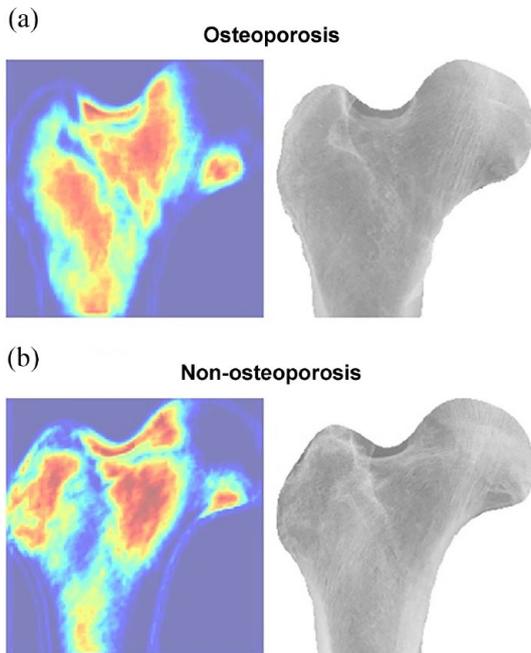
**Figure 3.** Relationship of the predicted *T*-score between DXA-measured and X1AI-Osteo-predicted *T*-score. The plot was created using the first imputed test set, which comprised 300 individuals. DXA, dual-energy X-ray absorptiometry.



**Figure 4.** Bland-Altman plots between DXA-measured and X1AI-Osteo-predicted *T*-score. Solid line, the mean difference (bias); upper and lower lines, 95% LoA. The mean difference was  $-0.010$  [95% CI:  $-0.019$  to  $-0.001$ ], LoA lower limit was  $-0.214$  [95% CI:  $-0.230$  to  $-0.198$ ], and upper limit was  $0.194$  [95% CI:  $0.178$ – $0.210$ ]. CI, confidence interval; DXA, dual-energy X-ray absorptiometry; LoA, limits of agreement.

further work applying the DL-based opportunistic osteoporosis screening using a simple hip radiograph. The several advantages of our DL-based screening tool include low cost, widely available radiographic modality, and simple protocol. Furthermore, it detects the risk of osteoporosis on the targeted hip, directly providing the future risk of hip fracture. For older people suffering a hip fracture, the concomitant screening of osteoporosis risk on the contralateral hip can be easily attained by X1AI-Osteo using the index hip radiograph, enabling clinicians to take early actions to prevent secondary fractures consistent with the spirit of Fracture Liaison Service.<sup>22</sup> Furthermore, owing to the simple protocol and high accuracy regarding DXA, the future work of X1AI-Osteo may be applied to the community osteoporosis screening in remote medical institutions or local clinics lacking DXA machines to optimize the strategy to expand screening populations cost-effectively.

Nevertheless, this study has some limitations. First, from opening the DICOM file to contour extraction and analysis, it takes approximately 12–15s (the DICOM file size is roughly 18–22MB). Second, the majority (60.4%) of our study population was diagnosed with osteoporosis, and approximately 14% of unilateral hip images were excluded owing to image-analyzing obstacles. Besides, the disease severity was relatively high because we enrolled patients from a tertiary medical center. Thus, our sample might not have represented the healthier aged population in the community. Third, all patients were enrolled from a single hospital. Our training datasets could be small such that overfitting of the CNN model might be a potential concern. Thus, the accuracy of DL-based osteoporosis classification and BMD prediction would be enhanced by increasing the number of images in a multicenter study. Besides, further external validation is warranted to confirm the applicability of our screen tool in other institutions. Finally, in our model of *T*-score prediction, we selected three clinical variables as additional features in the DL training metric to increase the prediction performance. Other confounding factors, including races, comorbidities, or previous fracture history, could be critical features contributing to the training performance of the CNN model. Nevertheless, the performance of our tool attained almost excellence in the BMD prediction even when only three clinical variables were included in the DL training metric.



**Figure 5.** Visualization of the region of interest by X1AI-Osteo on patients with and without osteoporosis. The relatively distinct areas of shading (obscure trabeculation) on the femoral neck and trochanteric region were identified as deep-learned feature areas.

## Conclusion

This study demonstrates that our proposed CNN model, X1AI-Osteo, may identify osteoporosis and predict  $T$ -scores on the targeted hip regions from simple hip radiographs with high accuracy. Hence, the future application of this screening tool could be an efficient strategy for population-based opportunistic osteoporosis screening with low cost and high adaptability for a broader population at risk. The previous model had a clinical intake of 300 people, and currently, it is being used by approximately 4000 people in the market.

## Declarations

### *Ethics approval and consent to participate*

The Ethics Committee of Taipei Medical University approved this study (registration number: TMU-JIRB N201909036). Owing to the retrospective nature of the study and the analysis of anonymous clinical data, the Ethics Committee waived the need for obtaining informed consent.

### *Consent for publication*

Not applicable.

### *Author contributions*

**Yu-Pin Chen:** Conceptualization; Investigation; Project administration; Resources; Validation; Writing – original draft; Writing – review & editing.

**Wing P. Chan:** Conceptualization; Methodology; Project administration; Resources; Supervision; Validation; Writing – review & editing.

**Han-Wei Zhang:** Conceptualization; Data curation; Formal analysis; Methodology; Software; Writing – original draft; Writing – review & editing.

**Zhi-Ren Tsai:** Conceptualization; Data curation; Formal analysis; Methodology; Software; Writing – original draft; Writing – review & editing.

**Hsiao-Ching Peng:** Formal analysis; Software; Visualization.

**Shu-Wei Huang:** Investigation; Validation.

**Yeu-Chai Jang:** Investigation; Validation.

**Yi-Jie Kuo:** Conceptualization; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.

### *Acknowledgements*

The authors are grateful to the programmer and data technologists from X1 Bone densitometer Solution and X1 Imaging for supporting this work, and did not receive any specific grant from external funding agencies.

### *Funding*

The authors express their gratitude to the Ministry of Science and Technology (grant numbers MOST 109-2314-B-038-037) and Wan Fang Hospital (grant numbers 112-wf-eva-20) for providing financial support for this research. This research was partially supported by the Ministry of Science and Technology through the Center for Precision Medicine Research of Asia University in Taiwan (grant number: NSTC 112-2321-B-468-001, NSTC 112-2410-H-468-008).

### *Competing interests*

The authors have read the journal's policy on disclosure of potential conflicts of interest and agreed to the journal's authorship statement. H-WZ and H-CP are employed by Biomedica Corporation and received support for using the products X1 Bone Densitometer Solution and X1 Imaging for conducting the present research. Both declare

that they had full access to the data in this study and take responsibility for the integrity and the accuracy of the analysis. No other author has reported potential conflict of interests relevant to this article, including relevant financial interests, activities, relationships, and affiliations.

#### Availability of data and materials

All data and related metadata underlying reported findings have been deposited in the public data repository: Mendeley Data (<https://data.mendeley.com>) with a digital object identifier (DOI) as <https://doi.org/10.17632/gmg3vvmvj4.1>.

#### Guarantor

The scientific guarantor of this publication is Yi-Jie Kuo.

#### Statistics and biometry

No complex statistical methods were necessary for this paper.

#### ORCID iDs

Yu-Pin Chen  <https://orcid.org/0000-0002-9729-6375>

Yi-Jie Kuo  <https://orcid.org/0000-0002-9889-5054>

#### Supplemental material

Supplemental material for this article is available online.

#### References

1. Ensrud KE and Crandall CJ. Osteoporosis. *Ann Intern Med* 2017; 167: ITC17–ITC32.
2. Chen YP, Kuo YJ, Liu CH, *et al.* Prognostic factors for 1-year functional outcome, quality of life, care demands, and mortality after surgery in Taiwanese geriatric patients with a hip fracture: a prospective cohort study. *Ther Adv Musculoskelet Dis* 2021; 13: 1759720X2111028360.
3. Chiang MH, Huang YY, Kuo YJ, *et al.* Prognostic factors for mortality, activity of daily living, and quality of life in Taiwanese older patients within 1 year following hip fracture surgery. *J Pers Med* 2022; 12: 102.
4. Chiang MH, Lee HJ, Kuo YJ, *et al.* Predictors of in-hospital mortality in older adults undergoing hip fracture surgery: a case–control study. *Geriatr Orthop Surg Rehabil* 2021; 12: 21514593211044644.
5. Blake GM and Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. *Postgrad Med J* 2007; 83: 509–517.
6. Abrahamsen B and Vestergaard P. Declining incidence of hip fractures and the extent of use of anti-osteoporotic therapy in Denmark 1997–2006. *Osteoporos Int* 2010; 21: 373–380.
7. Curtis JR, Laster A, Becker DJ, *et al.* The geographic availability and associated utilization of dual-energy X-ray absorptiometry (DXA) testing among older persons in the United States. *Osteoporos Int* 2009; 20: 1553–1561.
8. Singh M, Nagrath AR and Maini PS. Changes in trabecular pattern of the upper end of the femur as an index of osteoporosis. *J Bone Joint Surg Am* 1970; 52: 457–467.
9. Hauschild O, Ghanem N, Oberst M, *et al.* Evaluation of Singh index for assessment of osteoporosis using digital radiography. *Eur J Radiol* 2009; 71: 152–158.
10. Klatté TO, Vettorazzi E, Beckmann J, *et al.* The Singh index does not correlate with bone mineral density (BMD) measured with dual energy X-ray absorptiometry (DXA) or peripheral quantitative computed tomography (pQCT). *Arch Orthop Trauma Surg* 2015; 135: 645–650.
11. Yamamoto N, Sukegawa S, Kitamura A, *et al.* Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates. *Biomolecules* 2020; 10: 1534.
12. Jang R, Choi JH, Kim N, *et al.* Prediction of osteoporosis from simple hip radiography using deep learning algorithm. *Sci Rep* 2021; 11: 19997.
13. Yamamoto N, Sukegawa S, Yamashita K, *et al.* Effect of patient clinical variables in osteoporosis classification using hip X-rays in deep learning analysis. *Medicina (Kaunas)* 2021; 57: 846.
14. Hsieh CI, Zheng K, Lin C, *et al.* Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun* 2021; 12: 5472.
15. Kim S, Kim BR, Chae HD, *et al.* Deep radiomics-based approach to the diagnosis of osteoporosis using hip radiographs. *Radiol Artif Intell* 2022; 4: e210212.
16. von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577.
17. Chang YZ, Tsai ZR, Hwang JD, *et al.* Optimal fuzzy tracking control of uncertain nonlinear systems based on genetic algorithms and fuzzy

- Lyapunov function. *J Intell Fuzzy Syst* 2013; 24: 121–132.
18. Acharya T and Ray AK. *Image processing: principles and applications*. Wiley-Interscience, New York, NY, 2005.
  19. Russ JC. *The image processing handbook*. 4th ed. CRC Press, Inc., Subs. of Times Mirror 2000 Corporate Blvd. NW Boca Raton, FL, 2002.
  20. US Preventive Services Task Force; Curry SJ, Krist AH, Owens DK, *et al.* Screening for osteoporosis to prevent fractures: US Preventive Services Task Force recommendation statement. *JAMA* 2018; 319: 2521–2531.
  21. Doroudinia A and Colletti PM. Bone mineral measurements. *Clin Nucl Med* 2015; 40: 647–657; quiz p. 653–657.
  22. McLellan AR, Gallacher SJ, Fraser M, *et al.* The fracture liaison service: success of a program for the evaluation and management of patients with osteoporotic fracture. *Osteoporos Int* 2003; 14: 1028–1034.

Visit Sage journals online  
[journals.sagepub.com/  
home/tab](http://journals.sagepub.com/home/tab)

 Sage journals