Archives of Osteoporosis (2021) 16: 153
https://doi.org/10.1007/s11657-021-00985-8

ORIGINAL ARTICLE

Application of deep learning neural network in predicting bone mineral density from plain X-ray radiography

Chan-Shien Ho[1] · Yueh-Peng Chen[2,3] · Tzuo-Yau Fan[2] · Chang-Fu Kuo[2,4,5] · Tzu-Yun Yen[6,7] · Yuan-Chang Liu[8,9] · Yu-Cheng Pei[1,6,10,11]

Received: 27 November 2020 / Accepted: 2 August 2021 / Published online: 9 October 2021

Abstract

Summary DeepDXA is a deep learning model designed to infer bone mineral density data from plain pelvis X-ray, and it can achieve good predicted value for clinical use.

Purpose Osteoporosis is defined as a systemic disease of the bone characterized by a decrease in bone strength and deterioration of bone structure at the microscopic level, leading to bone fragility and increased risk of fracture. Bone mineral density (BMD) is the preferred method for the diagnosis of osteoporosis, and dual-energy x-ray absorptiometry (DXA) is the gold standard for diagnosing osteoporosis. Conventional radiography is more suited for the screening of osteoporosis rather than diagnosis, and osteoporosis can be detected on radiographs by experienced physicians only. This study explored the possibility of predicting BMD relative to DXA using patient radiographs.

Methods A deep learning algorithm of convolutional neural network (CNN) was used for the purpose. The method includes image segmentation, CNN learning, and a convolution-based regression model (DeepDXA) that links the isolated images of the femur bone to predict BMD value. Data were obtained in a single medical center from 2006 to 2018, with a total amount of 3472 pairs of pelvis X-ray and DXA examination within 1 year.

Results The proposed workflow successfully predicted BMD values of the femur bone with the correlation coefficient (R) of 0.85 ($P < 0.001$) and the accuracy of 0.88 for prediction osteoporosis, a finding that could be reliably ready for further clinical use.

Conclusion When suspicious osteoporosis is seen on plain films using the deep learning method we developed, further referral to DXA for the definite diagnosis of osteoporosis is indicated.

Keywords Osteoporosis · Bone mineral density · Convolutional neural network · Osteopenia · Dual-energy X-ray absorptiometry · Pelvis X-ray

Introduction

Osteoporosis is a systemic disease of the bone characterized by a decrease in bone strength and deterioration of bone microstructure, leading to bone fragility and increased risk of fracture [1]. Osteoporosis causes 1.5 million fractures per year in the USA, with a vast majority occurring in postmenopausal women. The prevalence of osteoporosis varies depending on whether it is defined by fracture incidence or low bone mineral density (BMD). For example, in the USA, there are approximately 40 million women with low BMD, but only 0.3 million hip fractures occur each year [2]. In Taiwan, according to data in 2009, 16,000 people suffer from hip fractures per year, and women are at twice the risk than men. The incidence of hip fracture rises rapidly with age (10% for hip fracture in Taiwanese women aged between 70 and 80 years) [3].

Bone strength is a composite performance that has several determinants, such as size, shape, architecture, and composition [4]. Bone strength is directional, such that the proximal femur bone strength is the greatest under axial loading consistent with habitual standing postures, but less than 50% for the impact direction induced by an incidental fall [5]. The bone composition is measured by BMD [6]. Although several methods, such as dual-energy X-ray absorptiometry (DXA), quantitative ultrasound (QUS), and

✉ Yueh-Peng Chen
yuepengc@gmail.com

✉ Yu-Cheng Pei
yspeii@gmail.com

Extended author information available on the last page of the article

photon absorptiometry have been developed, DXA is presently the gold standard diagnostic tool for osteoporosis [7, 8]. According to WHO (1994), the T-score (comparing a subject's BMD to that of a healthy 30-year-old female Caucasian), is used to diagnose osteoporosis. Normal T-score is defined as ≥ -1.0, osteopenia between -1.0 and -2.5, and osteoporosis < -2.5, which means that BMD of two and a half standard deviations below the mean for the reference-normal young adult [9]. Although first proposed for a subgroup of women, this method also applies to men [10, 11].

Plain X-ray are more suited for the screening of osteoporosis rather than diagnosis because there must be at least 30% to 50% of bone loss before the change is observable on a radiograph [12], a property that makes it difficult to rely on a physician's experiment to infer osteoporosis or osteopenia from plain X-ray. We can use semiquantitative method [13] to evaluate osteoporotic vertebral fracture, and the further diagnosis of osteoporosis should be performed using DXA. The prediction of BMD from conventional radiographs is difficult for both experienced physicians and present machine learning algorithms.

Methods utilizing radiographic parameters of the proximal or distal femur were developed to infer BMD and osteoporosis status. Nguyen et al. developed the cortical thickness index and canal flare index and showed that the prediction of osteoporosis using these indices had a sensitivity of 92.8% and specificity of 24.2% in female subjects [14]. He et al. used the mean cortical bone thickness and femoral cortex index of the distal femur to predict osteoporosis, showing an area under the receiver operating characteristic curve (AUROC) of 0.82 and 0.74, respectively [14, 15]. Clavert et al. found that the corticomedullar index of the distal humerus correlates with the BMD of the epiphysis with $R = 0.61$ [16]. For different targeted ROI, Navabi et al. also found that the mandibular cortical width derived from digital panoramic radiographs is significantly correlated with BMD assessed via DXA [17]. These findings suggested the utility of using plain X-rays to infer osteoporosis while showing a performance that is not satisfactory. The measurement and interpretation of BMD values with DXA only partly reflect bone strength, of which the regular arrangement of collagen, the degree of crosslinking of adjacent collagen fibrils, and mineral to protein matrix ratio may all contribute to bone quality. Some biomarkers of bone turnover have been reported to be predictive of fracture risk independent of measured BMD [18–20]. A variety of different methods were developed in an attempt to quantify those component determinants for bone strength. Methods in pursuit of presenting three-dimensional microarchitecture by two-dimensional DXA images [21–23] using the trabecular bone score, a measurement that analyzes the texture of DXA images [24] were proposed. Reports show that combined use of total BMD and trabecular bone score enhances the performance of fracture prediction, with AUROC for vertebral and hip fractures of 0.73 and 0.82, respectively [25, 26].

Due to the advancement in high-quality digitalized medical imaging and machine learning methods, computer-aided diagnostic systems play an important role in research and clinical settings in interpreting medical images. These systems have shown a giant leap of progress due to the utilization of a machine learning approach to image classification, of which convolutional neural network (CNN) is the most widely used. CNN is composed of functional layers for feature recognition of input images. First, the convolutional layers extract features from input images. Convolution is a mathematical operation taking inputs for calculation. Pooling layers reduce the number of parameters and downsize the input images. The last is a fully connected layer (FC) in which the neurons have connections to all activations in the previous layer. A growing number of studies have applied CNN on medical images, such as X-ray and computed tomography (CT), to predict bone mineral density or fracture risk [27].

The femur bones with higher T-scores (Supplementary Fig. 1a) have radiographic features of high radiopacity, smaller and fewer radiolucent pores, and thicker cortex than those with smaller T-scores (Supplementary Fig. 1b). Together with the differences in cortical thickness and bony architecture that reflect the contour and texture of the bony structures, respectively, the AI algorithm may be able to differentiate between images with and without osteoporosis. Given that plain X-ray images are frequently examined in the elderly, the methods to be developed in the present study aim to screen out high-risk patients so that fracture risk can be controlled at a lower cost, especially in developing countries. We also explored whether soft tissue or bony tissue that is outside of the femur bone was also used to infer BMD by comparing the performance of DeepDXA trained with background-included and background-removed images. We hypothesized that BMD values could be predicted from plain hip X-ray images without the need to define radiomic features, thus helping the physician to determine whether DXA should be arranged for each patient.

Material and methods

Data preparation

This study was approved by the Institutional Review Board of Chang Gung Medical Foundation (IRB approval number: 201801846B0). For deep learning model training, the input data were pelvic X-ray and the ground truth data were

BMD values obtained by DXA. Dataset was collected from historical images data of anteroposterior (AP) pelvic radiographs since January 2006 to December 2018 from a medical center (Chang Gung Memorial Hospital at Linkou) and 26,742 X-ray images and 19,726 DXA reports were first obtained. Among these X-ray images, we further selected those which received both X-ray and DXA done during this period of time and 17,633 X-ray images with 6347 corresponding DXA were further included. We selected a maximum 1-year period between which the X-ray and DXA were done and 9939 X-ray images and 3793 DXA reports were further included. If a subject had more than one X-ray radiograph-DXA pair, we kept the pair with the smallest time gap between the two examinations. Using this rule, 3793 input-ground truth pairs were selected, and from these pairs, 321 pairs were further excluded. The X-ray images with poor quality were excluded for deep learning model training. Specifically, in these X-ray images, 34 (10.59%) had hip fracture, 8 (2.49%) had shadow from medical devices, 9 (2.80%) had heterotopic ossificans around hip, 292 (90.96%) had poor positioning, and 39 (12.15%) had poor contrast, and 57 (17.76%) X-ray images had more than one of the aforementioned issues (Supplementary Fig. 2). Finally, 3472 input-ground truth pairs were used for further processing.

The DICOM files of the pelvic AP view images (Fig. 1a) were downloaded from the database with an average width*height resolution of 2497 ± 304 (pixels, mean $\pm$ S.D.) $\times 2645 \pm 227$ (pixels).

The BMD values for the total femur were retrieved from the DXA reports (Fig. 1b, green outline). The total femur BMD instead of portions of femur BMD (such as for neck, trochanteric, or intertrochanteric area) was used as we attempted to examine the model's ability to analyze a wider radiographic field. Almost all subjects had DXA of their left femur; their right and left femurs were assigned with the same BMD, assuming that bilateral femurs have comparable BMDs [19]. Finally, T-scores were transformed from BMD values using the following equation [28, 29]:

$$T - \text{score} = (\text{BMD} - 0.942)/0.122$$

Femur segmentation

Femur segmentation was performed using the BoneFinder® machine-learning software [30]. Using the generated polygonal annotations, pixel-wise masks that encompass the femur were obtained. Each annotation was reviewed by one physician, who was blind to the BMD value. Femurs with artificial implants were excluded. For each subject, the left and right femurs were assigned, yielding 2495 left and 2532 right femur contours.

Femur localization

Following femur segmentation, we further cropped the image by a square box along the outermost horizontal and vertical boundary of the femur contours. The edges of the cropped image flushed with the borders of the femur masks. The cropped image was denoted as the "background-included" image with a size of $(557 \pm 70) \times (679 \pm 91)$ pixels. The cropped image was segmented (using the annotations mentioned-above) such that the femur bone remains, and the non-bony parts such as the soft tissues and the background were removed (denoted as "background-removed" images). These background-removed images were masked with black color. The models were trained with background-included and background-removed images, respectively (Fig. 1c). We then compared the performance of DeepDXA trained with these two kinds of inputs.

Image augmentation and transformation

For data augmentation, each image was randomly rotated by an angle between $-20°$ and $20°$, scaled with a factor between 0.8 and 1.2, sheared with angles between $-15°$ and $15°$, vertically and horizontally translated with the proportion between -10% and 10%, and horizontally flipped with a 50% chance. Augmentation is performed to each training batch, for which its parameters were randomly assigned. Hence, following augmentation, the number of training images for each training batch was identical to that of the original training images. The augmented images were only used for the training process. Finally, each image was resized with the method of bilinear interpolation to 224×224 pixels. The intensity of each pixel of the DICOM file that ranged from 0 to 4095 was linearly transformed by dividing the intensity value by 4095 to yield a value that ranged from 0 to 1.

DeepDXA neural network development

We developed a convolutional neural network denoted as "DeepDXA" to predict BMDs from X-ray images. DeepDXA applied ResNet18 [31] (Fig. 1d). This architecture was chosen as it was more suitable for our regression tasks than deeper network architectures, such as ResNet50 [31], VGG16 [32], and DeepTEN [33]; the former focuses more on classifying the object's textures or microstructures instead of the object itself.

The initial weights were created by Kaiming normal initialization and the DeepDXA was separately trained with the background-included and removed images (Fig. 1c). For DeepDXA training, the inputs were cropped femur images of patients with at least one side of the hip without an artificial

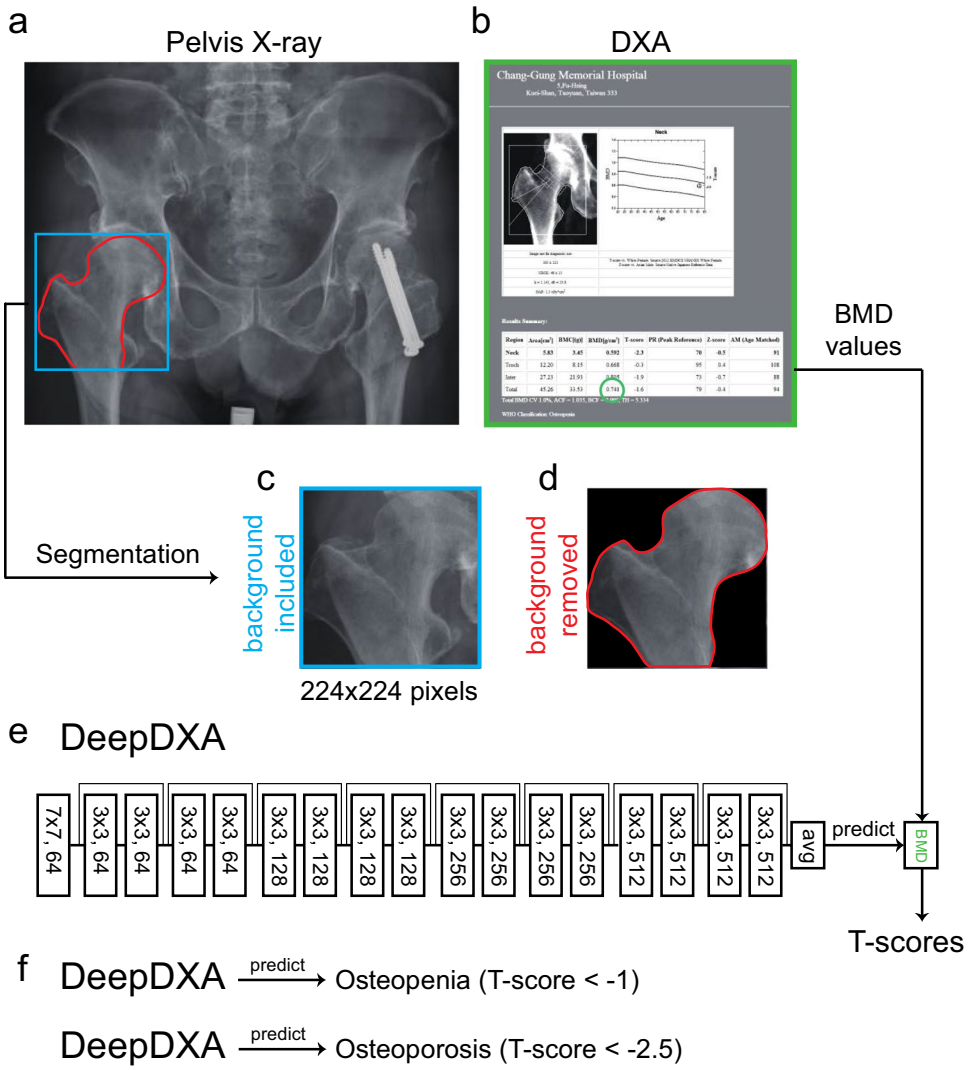

Fig. 1 Development of the DeepDXA. (**a**) Pelvis X-ray of the sample patient. The femur bone was segmented (red outline) and automatically cropped (cyan square) using algorithms. (**b**) DXA report of a patient. The "total BMD" value was retrieved from the report (green circle). We transformed BMD value to T-scores for the following analyses. (**c**) The DeepDXA model was trained with two sets of images: background-included and (**d**) background-removed images. (**e**) The architecture of the DeepDXA model. Numbers at different layers mean the convolutional filter-size and number of filters of each convolutional layer, e.g., (7×7, 64) means filter size 7×7 and 64 filters. (**f**) The DeepDXA was also trained to classify osteopenia and osteoporosis (see supplementary method)

implant (Fig. 1a), and the outputs were total BMDs. To modified the classification task of those models to the regression task, the fully connection layer as replaced with a single unit. The output of the unit, as an estimated value ($\hat{y}$), was used to evaluate the actual values (y). We used mean squared error (MSE) as our loss function, and the output of this layer was used to predict ground truth BMD.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$$

where n is the batch size of the training dataset.

Training procedures

Following the five-fold method, each image was randomly assigned to one of the five sets, which was cross-validated. If a subject had an X-ray taken several times, all the images were assigned to one of the training or validation datasets to avoid data leakage through images from identical patients. Each model was trained with randomly initialized weights. Each training session had 500 epochs, a batch size of 5, and a learning rate of 1e-4. Models were trained with Adam optimizer using NVIDIA 1080Ti GPU with 11 GB VRAM.

Metrics

During validation, MSE and Pearson's correlation coefficient (R) were used to evaluate the performance of the trained models. The distribution of the prediction errors ($y-\hat{y}$) of the model was analyzed by the error histogram and quantile–quantile plot (QQ plot). To examine the model's performance for binary classification, we use sensitivity as the ability of a test to correctly identify patients with osteoporosis/osteopenia and specificity as the ability of a test

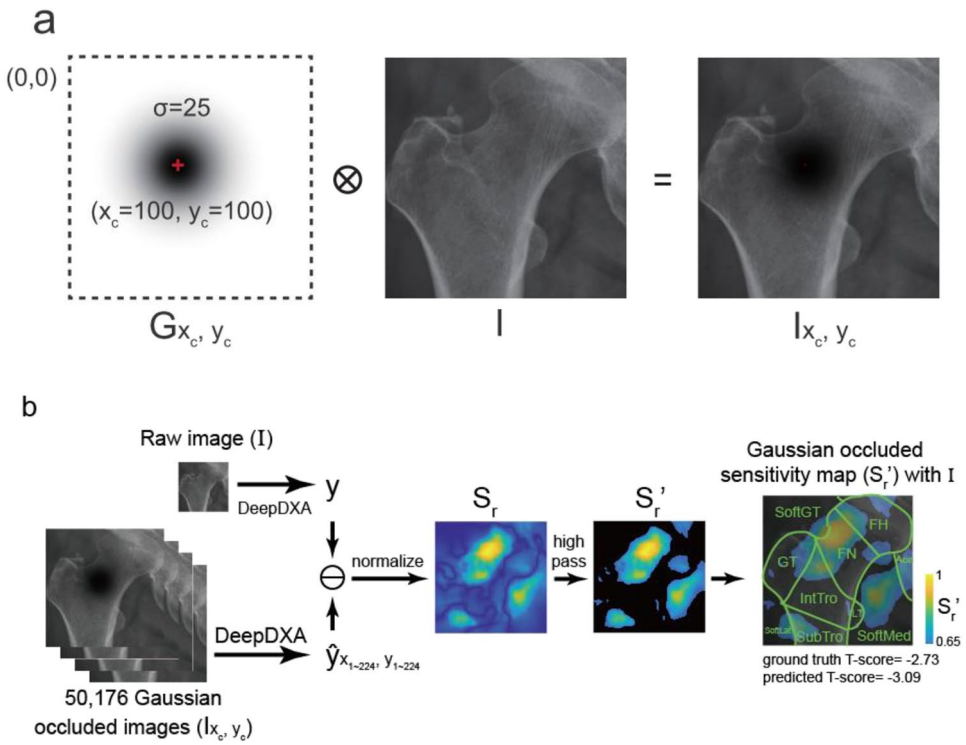

Fig. 2 The Gaussian occlusion sensitivity (GOS) map yielded by the analysis of Gaussian occlusion sensitivity. (**a**) The generation of the occlusion image (I_{x_c,y_c}) by the inner product of the original image (I) by the Gaussian mask (G_{x_c,y_c}) that centered as (x_c, y_c). (**b**) The generation of GOS maps. The predicted T-scores (y) from raw images were compared with the predicted T-score ($\hat{y}$) from Gaussian occluded images. The result was normalized (S_r) and we arbitrarily defined a high-pass threshold of 0.65 for S_r to exclude locations with minor S_r. S_r′ is the GOS map. We further superimposed the S_r′ with raw images (I). The example GOS map shows the S_r′ at 224 × 224 pixel-wise locations from the image (I). The hotspots are showed in the color of blue to yellow, with the yellow indicating the spots that have that maximal influence on model performance. The sample image (I) has ground truth and predicts a T-score of − 2.73 and − 3.09, respectively. FH, femoral head; FN, femoral neck; IntTro, intertrochanter; SubT, subtrochanter; GT, greater trochanter; BgH, background around femoral head; BgMed, background at medial side; BgGT, background around GT; BgLat, background at lateral side; Ace, acetabulum; LT, lesser trochanter; SoftMed, soft tissue at medial side; SoftGT, soft tissue around GT; SoftLat, soft tissue at lateral side

to correctly identify people without osteoporosis/osteopenia and the following equations were applied:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

where TP = correct prediction of positive osteoporosis/osteopenia status; FN = incorrect prediction of negative osteoporosis/osteopenia status; TN = correct prediction of negative osteoporosis/osteopenia status; FP = incorrect prediction of positive osteoporosis/osteopenia status.

Gaussian occlusion sensitivity

Occlusion sensitivity is a technique that infers the locations of the cues that account for the performance of a neural network by observing the degree to which the performance of inference of BMD is sabotaged by occluding a portion of the image using a Gaussian mask (Fig. 2a). To this end, for the validation images, each image was applied with 2D Gaussian masks following the equation:

$$G_{x_c,y_c}(x,y) = 1 - e^{-\left(\frac{(x-x_c)^2+(y-y_c)^2}{2\sigma^2}\right)}$$

where x_c and y_c define the location of the center of the Gaussian mask, and σ its spatial dispersion (σ=25 pixels). The occlusion image for a Gaussian mask (G_{x_c,y_c}), centered as (x_c, y_c), was computed using the following equation:

$$I_{x_c,y_c} = I \otimes G_{x_c,y_c}$$

where I_{x_c,y_c} and I denote the occlusion and original images, respectively, and ⊗ (Kronecker product) operation denotes pixel-by-pixel multiplication. Each of x_c and y_c ranges from 1 to 224 pixels, thus yielding 224 × 224 (a total of 50,176) occlusion images that cover the center locations of all possible Gaussian masks. The sensitivity (S_{x_c,y_c}) of each occlusion

location (x_c, y_c) reflects the degree to which the performance was biased by Gaussian occlusion following the equation:

$$S_{x_c,y_c} = |\hat{y}_{x_c,y_c} - y|$$

where $\hat{y}_{x_c,y_c}$ is the predicted T-score from the occlusion image I_{x_c,y_c}, and y is the predicted T-score from the original image I. Each original image thus yields 50,176 S_{x_c,y_c} as:

$$S = [S_{1,1}, S_{1,2}, \ldots, S_{224,224}]$$

where S is an ensemble of S_{x_c,y_c} among the whole Gaussian occlusion map. The heat map that represents the normalized relative gaussian occlusion sensitivity for each original image was computed as:

$$S_r = \frac{S - \min(S)}{\max(S) - \min(S)}$$

where S_r represents the normalized sensitivity with its minimum and maximum of 0 and 1, respectively. We arbitrarily defined a high-pass threshold of 0.65 for S_r to exclude locations with minor S_r, such that:

$$S_r = \begin{cases} S_r, if S_r \geq 0.65 \\ 0, if S_r < 0.65 \end{cases}$$

where S_r is the gaussian occlusion sensitivity (GOS) map (Fig. 2b).

To annotate the location of the hot spots in the GOS maps, the GOS maps with background removed were divided by 9 locations (Supplementary Fig. 3a), and those with background included were divided by 10 locations (Supplementary Fig. 3b). The difference of the annotation system was accounted for by the fact that the acetabulum and the lesser trochanter were not included in the images with the background removed. The distribution of the hot spots across these areas was manually counted for each image. The analysis included data obtained from 150 randomly sampled subjects consisting of 50 patients with osteoporosis, 50 with osteopenia, and 50 normal subjects. We specifically identified the locations whose bias of sensitivity > 0.825, an arbitrary threshold that we considered a reflection of a substantial influence on DeepDXA performance.

Statistics

All predicted BMD values were first transformed into T-scores [28]. Pearson's correlation between the ground truth (y) and predicted T-scores ($\hat{y}$) was used to analyze the models' performance for the validation dataset. The Student's t-test was used to compare data characteristics between male and female subjects. The models' performance was reflected by the correlation coefficients (R) and MSE values for predicting the validation dataset. Wilcoxon rank-sum test was used to compare the correlation coefficient and MSE of DeepDXA model when DeepDXA was separately trained with the background-included and removed images. The normality test was performed using the QQ plot with a one-sample Kolmogorov–Smirnov test. One-way ANOVA or paired t-test was applied to compare the number of hot spots detected in the GOS maps among the normal, osteopenia, and osteoporosis groups (Supplementary Table 2). A paired t-test was used to compare the number of hot spots between the background-included and background-removed conditions (Supplementary Table 3). For each location, the Chi-squared test or Fisher's exact test was used to compare the probability of having hot spots among the three groups (Supplementary Table 4).

Table 1 Data characteristics

	All	Male	Female	*P* value
Age (years)	71.2 ± 13	71.1 ± 14.3	71.2 ± 12.6	*P = 0.86*
Image number (*n*)	2800	609	2191	
Left hip (*n*)	2495	525	1970	
Right hip (*n*)	2532	590	1942	
Any hip (*n*)	5027	1115	3912	
Total BMD (g/cm^2)	0.66 ± 1.57	0.73 ± 0.16	0.63 ± 0.15	**P < 0.001
T-score	−2.3 ± 1.29	−1.73 ± 1.27	−2.5 ± 1.25	**P < 0.001
Left hip DXA (*n*)	2037	393	1644	
Right hip DXA (*n*)	763	216	547	
Time gap† (day)	65.7 ± 92.8	56.6 ± 88.1	68.3 ± 93.9	*P = 0.006
Energy (kVp)	77.9 ± 7.9	-	-	
Intensity (mAs)	33.3 ± 19.8	-	-	
Distance of source to detector (mm)	1012.1 ± 54.4	-	-	
Exposure time (ms)	119.1 ± 68.7	-	-	

†Absolute time gap between X-ray and DXA

Data were presented as mean ± standard deviation

*$P < 0.01$, **$P < 0.001$, Student's t-test

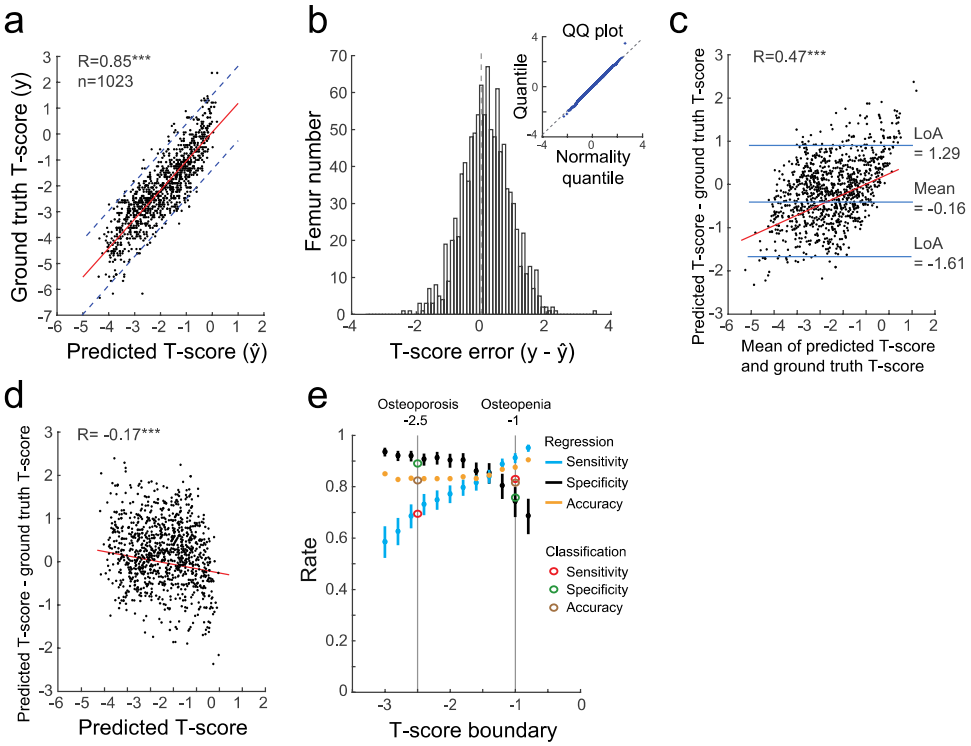

Fig. 3 Performance of DeepDXA for predicting T-scores for images with background removed. (**a**) Ground truth T-score as a function of predicted T-score showed a significant linear correlation ($R = 0.85$, $P < 0.001$). The red line represents the regression line:ground truth T-score = 1.12 × (predicted T-score) + 0.07, and the blue dotted lines represent the 95% prediction bounds. (**b**) The histogram of the prediction error for the T-score. The QQ plot indicates that the prediction error followed a normal distribution. (**c**) The Bland–Altman plot of DXA vs DeepDXA. (predicted T-score − ground truth T-score) = 0.29 × (mean of predicted and ground truth T-score) + 0.43 ($R = 0.465$, $P < 0.001$). LoA, limits of agreement. (**d**) The correlation between the difference and predicted T-score. (predicted − ground truth T-score) = −0.19 × (predicted T − score) − 0.069 ($R = -0.171$, $P < 0.001$). (**e**) The sensitivity and specificity of DeepDXA for predicting whether a subject's T-score is less than a cut-point criterion. The left and right gray lines indicate the criteria for osteoporosis and osteopenia, respectively. The sensitivity, specificity, and accuracy for the prediction of regression DeepDXA model were shown in cyan, black, and yellow dots respectively. The DeepDXA showed a sensitivity of 0.94, a specificity of 0.65, and an accuracy of 0.88 for osteopenia (T-score criteria = −1.0), and a sensitivity of 0.76, a specificity of 0.87, and an accuracy of 0.84 for osteoporosis (T-score criteria = −2.5). The sensitivity, specificity, and accuracy for the prediction of classification DeepDXA model were shown in red, green, and brown dots, respectively. For the classification DeepDXA model, please refer to Supplementary Method and Supplementary Fig. 6and Supplementary Table 1. ***$P < 0.001$, Pearson's correlation

Results

Data characteristics

A total of 2800 (male = 609, female = 2191) pelvis X-ray plain films were included (Table 1), providing 5027 unilateral femur images (male = 1115, female = 3912). Compared with male subjects, females had significantly lower total BMD values (Student's *t*-test, $P < 0.001$) and T-scores ($P < 0.001$). Male and female subjects had comparable age (male: 71.1 ± 14.3 years, female: 71.2 ± 12.6 years, Student's *t*-test, $P = 0.86$). The mean kVp value of training images was 77.9 ± 7.9, the mean mAs value was 33.4 ± 19.8, the mean of distance of source to detector was 1012.1 ± 54.4 (mm), and the mean exposure time was 119.1 ± 68.7(ms). As demographic data showed in Table 1, the actual time gaps between X-ray and DXA were 65.7 ± 92.8 and 68.3 ± 93.9 days in male and female subjects respectively, with most of the time gaps clustered within 6 months.

Table 2 Performance of the two separately trained DeepDXA models: background-included and background-removed

Performance/condition	Background-included	Background-removed	*P* value
R value	0.84 ± 0.02	0.81 ± 0.03	0.22
MSE	0.0091 ± 0.001	0.0095 ± 0.0007	0.55

R correlation coefficient, *MSE* mean squared error

Data were presented as mean ± standard deviation

*$P < 0.05$, Wilcoxon rank-sum test

DeepDXA neural network

For each test during the fivefold validation, DeepDXA was trained and then validated with 3972 and 1055 images, respectively. Figure 3a illustrates the prediction of T-scores (linearly transformed from predicted BMDs in METHODS) of DXA from the input images, showing a promising correlation with a coefficient (R) of 0.85 ($P<0.001$, $R^2=0.73$, slope = 1.12, t value = 52.76, df = 1041). We also compared the performance using DeepDXA on DXA-Xray pairs with time gap > 180 days with ≦180 days (Supplementary Fig. 4). The results showed that DeepDXA using data (hip images $n=831$) with time gap > 180 days showed a correlation coefficient of 0.833 ($P<0.001$) while DeepDXA using data (hip images $n=4196$) with time gap ≦180 days showed a correlation coefficient of 0.828 ($P<0.001$). There was no difference of correlation coefficients between these two groups ($P=0.72$, two sample t-test). Accuracy of prediction of osteoporosis data was 84.5% and 83.8% in the time gap > 180 and ≦180 days groups, respectively, a finding suggesting that these two groups did not differ in the performance of the prediction of osteoporosis. The residual, the error T-score estimation ($y-\hat{y}$), followed a normal distribution (Fig. 3b, 0.16 ± 0.74) (KS test, $P=0.91$), and was uncorrelated to ground-truth validation data (y) (Spearman's correlation, $R^2=0.0005$, $P=0.46$), indicating that DeepDXA did not yield a systematic error. We can observe a good match between the predicted and the ground truth T-score values in the Bland–Altman plot [34] (Fig. 3c). There is a positive correlation (linear regression: (predicted T − score − ground truth T − score) $= 0.29 \times$ (mean of predicted and ground truth T − score)+ $0.43, R = 0.465, P < 0.001$) between the difference between predicted and ground truth T-score and the mean of ground truth and predicted T-score in the Bland–Altman plot, indicating the existence of a slight proportional bias. This bias could be accounted for by predicting errors for extreme T-score values. This systematic bias can also be assessed by analyzing the correlation between the difference between predicted and ground truth T-score and predicted T-score. A weak negative correlation ($R=-0.17$, coefficient $=-0.012$) was observed. This correlation is not observed if we delimit predicted T-score between − 3.5 and − 0.5 and plot the difference between predicted and ground truth T-score plotted against predicted T-score between − 3.5 and − 0.5 (Supplementary Fig. 5), indicating less systematic prediction bias in this range of predicted T-scores. Finally, we examined the performance of DeepDXA for predicting whether a subject had osteoporosis or osteopenia (Fig. 3e). The DeepDXA showed a sensitivity of 0.94, a specificity of 0.65, and an accuracy of 0.88 for osteopenia (T-score criteria = − 1.0), and a sensitivity of 0.76, a specificity of 0.87, and an accuracy of 0.84 for osteoporosis (T-score criteria = − 2.5), supporting its utility to infer osteoporosis or osteopenia. For the classification DeepDXA model, please refer to Supplementary Method and Supplementary Fig. 6 and Supplementary Table 1.

Across the five validation tests, correlation coefficients (R) were significant (all $P<0.001$) both in images with background-included ($R=0.84\pm0.02$) or removed ($R=0.81\pm0.03$) conditions (Table 2), and there were no differences between the two conditions in their R-value or (Wilcoxon rank-sum test, $P=0.22$) MSE (Wilcoxon rank-sum test, $P=0.55$). When soft tissue was included in input images for model training, the specificity for detection of osteoporosis was significantly higher (0.90 ± 0.04) than that with background-removed images (0.82 ± 0.03, $P=0.03$) (Table 3).

Gaussian occlusion sensitivity (GOS), which determines which image features were used to infer total BMD, showed that the locations of the hot spots on the GOS maps were different for background-included and removed conditions (Supplementary Tables 2–4). In the background-included condition, the number of locations of GOS hot spots increased from 5.06 ± 1.53 in the normal group to 7.48 ± 1.37 in the osteopenia group and 8.56 ± 1.20 in the osteoporosis groups (ANOVA, $P<0.001$) (Supplementary Table 2). Similarly, in the background removed condition, the number of locations of GOS hot spots also increased from 5.06 ± 1.15 in the normal group to 6.86 ± 1.31 and 6.62 ± 1.31 in the osteopenia and osteoporosis groups, respectively (ANOVA, $P<0.001$). Specifically, the hot spots of both background-included and removed groups were present in FN, GT, IntTro, and BgMed/SoftMed, but the background-included group showed more hot spots in SubTro and BgLat (Supplementary Table 3). These results indicate that the locations of femur bone and areas of soft tissue around the femur were both important for inferring T-scores.

The GOS map also differed across the low, middle, and high T-score groups (Supplementary Fig. 7). GOS maps showed a wide spread of hot spots in the low T-score group, but a concentration of hot spots in FN and BgMed/SoftMed, the high T-score group. The middle T-score group showed an intermediate finding. These results indicate that, when the features of bone quality presented by X-ray images of bony structures themselves can provide sufficient information for DeepDXA to infer T-scores, the information from the soft tissue background is not necessarily used by DeepDXA to infer T-scores.

Discussion

To the best of our knowledge, the present study is the first to predict bone density from hip X-ray images using the CNN regression model. The results showed high sensitivity and specificity for identifying osteoporosis, suggesting its clinical utility. The hot spots in the GOS maps showed that

Table 3 Performance of the two separately trained DeepDXA models: background-included and background-removed

Performance/condition		Background-included	Background-removed	*P* value
Osteoporosis	Sensitivity	0.73 ± 0.05	0.80 ± 0.04	$P = 0.06$
	Specificity	0.90 ± 0.04	0.82 ± 0.03	$^*P = 0.03$
Osteopenia	Sensitivity	0.93 ± 0.03	0.95 ± 0.02	$P = 0.31$
	Specificity	0.63 ± 0.09	0.53 ± 0.1	$P = 0.1$

Data were presented as mean ± standard deviation

$^*P < 0.05$, Wilcoxon rank-sum test

the image location used to infer BMD was different across different BMD levels. Notably, although the femoral neck is mostly used, other areas such as the lesser trochanter, greater trochanter, intertrochanter, and even nearby soft tissues showed substantial importance. The inclusion of the trochanteric area is compatible with the previous findings that the bony cortex is important, suggesting that the micro-architecture reflected in the texture pattern on X-ray images may also contribute to the prediction of BMD. This observation supported our hypothesis that an X-ray image contains a variety of information, such as bone size, shape, and micro-architecture that can be used to infer bone strength. Interestingly, the inclusion of soft tissue around the femur did not improve the performance.

Traditional machine learning methods using radiomics as cues have been developed to predict the presence of osteoporosis or osteopenia in patients. Rastegar et al. used radiomics of seven regions, such as lumbar, vertebral, and femoral segments, to predict osteoporosis and yield AUROC of 0.50 to 0.78. The advantage of CNN is that rule-based processed or manual pre-processing is no longer needed by establishing an automatic pipeline of processing. Tecle et al. used X-ray radiography of the central third of the second metacarpal bone to predict osteoporosis and achieved a sensitivity of 82.4% and specificity of 94.3%, differentiating osteoporotic from non-osteoporotic subjects [35]. Yasaka et al. used CNN to show that the BMD of lumbar vertebrae can be estimated from unenhanced abdominal CT images, and the estimated values correlate with the BMD values obtained by DXA. Osteoporosis was diagnosed with AUROC of 0.970, which showed a significant correlation with the BMD values obtained with DXA [36]. Liu et al. used U-Net to predict osteoporosis from pelvic X-ray radiography based on data collected from 89 subjects and showed an accuracy for differentiating osteopenia from osteoporosis with a sensitivity of 61% and specificity of 63% [37]. Lee et al. proposed a combination of feature extraction from spine X-ray images of a Korean population using VGGnet and classification by random forest, yielding a sensitivity of 0.81 and specificity of 0.60 to identify osteopenia.

In the present study, the DeepDXA reached a sensitivity of 0.94 and specificity of 0.65 for osteopenia and a sensitivity of 0.76 and specificity of 0.87 for osteoporosis. The improvement of specificity compared with previous models could be attributed to the fact that DeepDXA is based on ResNet18 [31], which may be superior in texture-based analysis for medical images [38, 39]. Moreover, our cropping method, including femur bone segmentation, may account for a better outcome. The performance of DeepDXA suggests its potential to meet clinical needs by inferring the risk of osteoporosis and osteopenia so that physicians can arrange further evaluations to provide early intervention and ultimately prevent possible osteoporotic fractures, an approach that can assist the patients and reduce the burden of medical care.

However, this model had several potential limitations. First, the specific aim of the present study is to illustrate the performance of deep learning on real-world data. As a retrospective study that analyzed real-world data, the setting of the X-ray, such as energy, intensity, distance of source to detector, and exposure time indeed varied widely across subjects. Plain films were not performed on the same X-ray machine since there are different examination stations in our medical center and the parameters were set by different examiners to optimize its diagnostic value. Interestingly, using images from different acquisition parameters as training data, we could avoid the model to be overfitted and the trained model may have generalization capability over different machines and subjects. Second, the real-world data has more femur X-ray radiography than DXA data, with female dominance. Also, in our data, females had lower T-scores. Females are more prone to postmenopausal osteoporosis, while males have a higher probability of developing idiopathic osteoporosis [40, 41]. Given the difference in pathophysiology between genders, the future model may apply a different network for males and females. Additionally, most BMD was examined for the left femur, but in the present study, we used both the left and right femur to predict the BMD of the left femur, an approach that could cause more prediction errors.

Third, the present study did not include demographic data as input variables of DeepDXA. Factors such as body height

and weight may affect bone size and areal BMD [42]. BMD and X-ray radiography images are a result of projection from three-dimensional X-ray absorption to a two-dimensional image. Therefore, a bigger bone size will have higher intensity in the X-ray radiography [4, 43]. Age, comorbidity, body height, and body weight may also contribute to bone quality [44], such that the inclusion of these parameters may improve the performance of DeepDXA. Fourth, the present study is based on a population of patients with a mean age of 71 years, and thus, future studies are needed to improve this model so that the prediction can be extrapolated to a wide age range population. Fifth, the period between X-ray and DXA was limited to be 1 year in the present study. Elderly women with a hip fracture were shown to have a decreased in BMD at femoral neck of 2.1%, 2.5%, and 4.6%, and intertrochanter of 0.7%, 1.4%, and 2.1% at 2, 6, and 12-month follow-up, respectively [45]. Therefore, some subjects that suffered from hip fracture and immobilization due to surgery might have a change of BMD during this 1-year period. However, the demographic data showed that most of the time gap between X-ray and DXA were clustered within 6 months, suggesting that the effect of postoperative bone loss might only have a minor influence on our results. Sixth, a lack of sample size with T-score with extreme values (extremely high/low T-scores) in the training dataset can result in a decreased performance of bone density prediction. The best way to improve the performance for extreme T-score values is to increase training dataset in this extreme value range. As a real-word data that was collected for 10 years from a medical center, we think of that this imbalance just reflects the patient population that seeks bone density evaluation. It is suggested that a prospect data that intentionally collect extreme data might be needed. For clinical use, we are more confident with the prediction only when the predicted T-scores fall between − 3.5 and − 0.5. Finally, traditional DXA analysis can infer BMD by analyzing images from different parts of the femur bone, such as the femoral neck, trochanter, intertrochanter, ward triangle, and total hip image. Future studies are needed to examine whether DeepDXA can robustly infer BMD of a specific part of the femur, such as the intertrochanter. Also, a correlation of 0.85 or so with DXA BMD is not great but this may not be so important for fracture prediction. The specific aim of the present study is to use hip X-ray to infer BMD, from which the decision on further referral to DXA study could be made by the physician. We agree that the ultimate goal of the deep learning model is to predict fracture risk, the information that is important for clinical decision making. One of our present lines of inquiry is to examine the value of deep learning for the prediction of fracture. A hybrid model that incorporates X-ray images and patient characteristics might be tested in the future.

Conclusions

The deep learning model proposed in this study has good performance characteristics for predicting osteoporosis from hip X-ray images, enabling early detection of osteopenia and osteoporosis at a low cost. Future studies will examine the clinical utility of deep learning models by evaluating the hybrid method that includes BMD and X-ray images for the prediction of osteoporotic fracture.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11657-021-00985-8.

Acknowledgements The authors thank the statistical assistance and wish to acknowledge the support of the Maintenance Project of the Center for Artificial Intelligence in Medicine at Chang Gung Memorial Hospital (Grant CLRPG3H0012, CIRPG3H0012) for study design and monitor, data analysis, and interpretation and Chang Gung Medical Foundation Grant (CMRPG5H0051-3, CMRPG3K0231-2) for manpower and data analysis.

Author contribution C.-S.H., Y.-P.C., T.-Y.F., C.-F.K., and Y.-C.P. designed research; C.-S.H., Y.-P.C., T.-Y.F., and T.-T.Y. collected data; C.-S.H., Y.-P.C., T.-Y.F., C.-F.K., and Y.-C.P. analyzed data; C.-S.H., Y.-P.C., Y.-C.L, and Y.-C.P. wrote the paper.

Declarations

Conflicts of interest None.

References

1. Consensus A (1993) Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. Am J Med 94(6):646–650
2. Solomon C, Black D, Rosen C (2016) Postmenopausal osteoporosis. N Engl J Med 374(3):254–262
3. Shao C-J, Hsieh Y-H, Tsai C-H, Lai K-A (2009) A nationwide seven-year trend of hip fractures in the elderly population of Taiwan. Bone 44(1):125–129
4. Choksi P, Jepsen KJ, Clines GA (2018) The challenges of diagnosing osteoporosis and the limitations of currently available tools. Clinical Diabetes and Endocrinology 4(1):12. https://doi.org/10.1186/s40842-018-0062-7
5. Keyak JH, Skinner HB, Fleming JA (2001) Effect of force direction on femoral fracture load for two types of loading conditions. J Orthop Res 19(4):539–544
6. Prevention O (2000) Diagnosis, and therapy. NIH Consens Statement 17(1):1–36
7. Kanis JA (1994) Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. Osteoporos Int 4(6):368–381
8. Kanis JA (2002) Diagnosis of osteoporosis and assessment of fracture risk. The Lancet 359(9321):1929–1936
9. Kanis JA, Melton LJ III, Christiansen C, Johnston CC, Khaltaev N (1994) The diagnosis of osteoporosis. J Bone Miner Res 9(8):1137–1141
10. Baim S, Binkley N, Bilezikian JP, Kendler DL, Hans DB, Lewiecki EM, Silverman S (2008) Official positions of the

International Society for Clinical Densitometry and executive summary of the 2007 ISCD Position Development Conference. J Clin Densitom 11(1):75–91
11. Schousboe JT, Shepherd JA, Bilezikian JP, Baim S (2013) Executive summary of the 2013 International Society for Clinical Densitometry Position Development Conference on bone densitometry. J Clin Densitom 16 (4):455–466. doi:https://doi.org/10.1016/j.jocd.2013.08.004
12. Anil G, Guglielmi G, Peh WC (2010) Radiology of osteoporosis. Radiol Clin North Am 48(3):497–518. https://doi.org/10.1016/j.rcl.2010.02.016
13. Genant HK, Wu CY, Van Kuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 8(9):1137–1148
14. Nguyen BNT, Hoshino H, Togawa D, Matsuyama Y (2018) Cortical thickness index of the proximal femur: a radiographic parameter for preliminary assessment of bone mineral density and osteoporosis status in the age 50 years and over population. Clin Orthop Surg 10(2):149–156
15. He Q, Sun H, Shu L, Zhu Y, Xie X, Zhan Y, Luo C (2018) Radiographic predictors for bone mineral loss: cortical thickness and index of the distal femur. Bone & joint research 7(7):468–475
16. Clavert P, Javier R-M, Charrissoux J, Obert L, Pidhorz L, Sirveaux F, Mansat P, Fabre T (2016) How to determine the bone mineral density of the distal humerus with radiographic tools? Surg Radiol Anat 38(4):389–393
17. Samelson EJ, Broe KE, Xu H, Yang L, Boyd S, Biver E, Szulc P, Adachi J, Amin S, Atkinson E, Berger C, Burt L, Chapurlat R, Chevalley T, Ferrari S, Goltzman D, Hanley DA, Hannan MT, Khosla S, Liu C-T, Lorentzon M, Mellstrom D, Merle B, Nethander M, Rizzoli R, Sornay-Rendu E, Van Rietbergen B, Sundh D, Wong AKO, Ohlsson C, Demissie S, Kiel DP, Bouxsein ML (2019) Cortical and trabecular bone microarchitecture as an independent predictor of incident fracture risk in older women and men in the Bone Microarchitecture International Consortium (BoMIC): a prospective study. Lancet Diabetes Endocrinol 7(1):34–43. https://doi.org/10.1016/S2213-8587(18)30308-5
18. Vasikaran S, Eastell R, Bruyère O, Foldes A, Garnero P, Griesmacher A, McClung M, Morris HA, Silverman S, Trenti T (2011) Markers of bone turnover for the prediction of fracture risk and monitoring of osteoporosis treatment: a need for international reference standards. Osteoporos Int 22(2):391–420
19. Johnell O, Odén A, De Laet C, Garnero P, Delmas P, Kanis J (2002) Biochemical indices of bone turnover and the assessment of fracture probability. Osteoporos Int 13(7):523
20. Garnero P, Sornay-Rendu E, Duboeuf F, Delmas PD (1999) Markers of bone turnover predict postmenopausal forearm bone loss over 4 years: the OFELY study. J Bone Miner Res 14(9):1614–1621
21. Silva BC, Leslie WD, Resch H, Lamy O, Lesnyak O, Binkley N, McCloskey EV, Kanis JA, Bilezikian JP (2014) Trabecular bone score: a noninvasive analytical method based upon the DXA image. J Bone Miner Res 29(3):518–530
22. Winzenrieth R, Michelet F, Hans D (2013) Three-dimensional (3D) microarchitecture correlations with 2D projection image gray-level variations assessed by trabecular bone score using high-resolution computed tomographic acquisitions: effects of resolution and noise. J Clin Densitom 16(3):287–296
23. Harvey NC, Glüer CC, Binkley N, McCloskey EV, Brandi ML, Cooper C, Kendler D, Lamy O, Laslop A, Camargos BM, Reginster JY, Rizzoli R, Kanis JA (2015) Trabecular bone score (TBS) as a new complementary approach for osteoporosis evaluation in clinical practice. Bone 78:216–224. https://doi.org/10.1016/j.bone.2015.05.016
24. Hans D, Barthe N, Boutroy S, Pothuaud L, Winzenrieth R, Krieg M-A (2011) Correlations between trabecular bone score, measured using anteroposterior dual-energy X-ray absorptiometry acquisition, and 3-dimensional parameters of bone microarchitecture: an experimental study on human cadaver vertebrae. J Clin Densitom 14(3):302–312. https://doi.org/10.1016/j.jocd.2011.05.005
25. Martineau P, Silva BC, Leslie WD (2017) Utility of trabecular bone score in the evaluation of osteoporosis. Curr Opin Endocrinol Diabetes Obes 24(6):402–410. https://doi.org/10.1097/med.0000000000000365
26. Silva BC, Leslie WD (2017) Trabecular bone score: a new DXA–derived measurement for fracture risk assessment. Endocrinol Metab Clin 46(1):153–180
27. Derkatch S, Kirby C, Kimelman D, Jozani MJ, Davidson JM, Leslie WD (2019) Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: a registry-based cohort study of dual X-ray absorptiometry. Radiology 293(2):405–411
28. Melamed A, Vittinghoff E, Sriram U, Schwartz AV, Kanaya AM (2010) BMD reference standards among South Asians in the United States. J Clin Densitom 13(4):379–384. https://doi.org/10.1016/j.jocd.2010.05.007
29. Cundy T, Cornish J, Evans MC, Gamble G, Stapleton J, Reid IR (1995) Sources of interracial variation in bone mineral density. J Bone Miner Res 10(3):368–373
30. Lindner C, Thiagarajah S, Wilkinson JM, Wallis GA, Cootes TF, Consortium a (2013) Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE transactions on medical imaging 32 (8):1462-1472
31. He K, Zhang X, Ren S, Sun J Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. pp 770–778
32. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556
33. Zhang H, Xue J, Dana K Deep ten: Texture encoding network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. pp 708–717
34. Yu M, Tham Y-C, Rim TH, Ting DSW, Wong TY, Cheng C-Y (2019) Reporting on deep learning algorithms in health care. The Lancet Digital Health 1(7):e328–e329. https://doi.org/10.1016/S2589-7500(19)30132-3
35. Tecle N, Teitel J, Morris MR, Sani N, Mitten D, Hammert WC (2020) Convolutional neural network for second metacarpal radiographic osteoporosis screening. The Journal of Hand Surgery. https://doi.org/10.1016/j.jhsa.2019.11.019
36. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O (2020) Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. Eur Radiol. https://doi.org/10.1007/s00330-020-06677-0
37. Liu J, Wang J, Ruan W, Lin C, Chen D (2019) Diagnostic and gradation model of osteoporosis based on improved deep U-Net network. J Med Syst 44(1):15. https://doi.org/10.1007/s10916-019-1502-3
38. Gay J, Harlin H (2019) Texture-based classification for oral cancer detection: implementation and performance analysis of deep learning approaches. Networks (RotEqNet) 8:11
39. Hu J, Song W, Zhang W, Zhao Y, Yilmaz A (2019) Deep learning for use in lumber classification tasks. Wood Sci Technol 53(2):505–517
40. Riggs BL, Melton LJ, Robb RA, Camp JJ, Atkinson EJ, McDaniel L, Amin S, Rouleau PA, Khosla S (2008) A population-based assessment of rates of bone loss at multiple skeletal sites: evidence for substantial trabecular bone loss in young adult women and men. J Bone Miner Res 23(2):205–214. https://doi.org/10.1359/jbmr.071020

41. Herrera A, Lobo-Escolar A, Mateo J, Gil J, Ibarz E, Gracia L (2012) Male osteoporosis: a review. World J Orthop 3(12):223–234. https://doi.org/10.5312/wjo.v3.i12.223
42. Wang L, Ran L, Zha X, Zhao K, Yang Y, Shuang Q, Liu Y, Hind K, Cheng X, Blake GM (2020) Adjustment of DXA BMD measurements for anthropometric factors and its impact on the diagnosis of osteoporosis. Arch Osteoporos 15(1):155. https://doi.org/10.1007/s11657-020-00833-1
43. Lochmüller EM, Miller P, Bürklein D, Wehr U, Rambeck W, Eckstein F (2000) In situ femoral dual-energy X-ray absorptiometry related to ash weight, bone size and density, and its relationship with mechanical failure loads of the proximal femur. Osteoporos Int 11(4):361–367. https://doi.org/10.1007/s001980070126
44. Boskey AL, Imbert L (2017) Bone quality changes associated with aging and disease: a review. Ann N Y Acad Sci 1410(1):93–106. https://doi.org/10.1111/nyas.13572
45. Fox KM, Magaziner J, Hawkes WG, Yu-Yahiro J, Hebel JR, Zimmerman SI, Holder L, Michael R (2000) Loss of bone density and lean body mass after hip fracture. Osteoporos Int 11(1):31–35. https://doi.org/10.1007/s001980050003

Authors and Affiliations

Chan-Shien Ho[1] · Yueh-Peng Chen[2,3] · Tzuo-Yau Fan[2] · Chang-Fu Kuo[2,4,5] · Tzu-Yun Yen[6,7] · Yuan-Chang Liu[8,9] · Yu-Cheng Pei[1,6,10,11]

1 Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital at Linkou, No. 5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan

2 Center for Artificial Intelligence in Medicine, Chang Gung Memorial Hospital at Linkou, No. 5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan

3 Department of Industrial Design, College of Management, Chang Gung University, Taoyuan, Taiwan

4 Division of Rheumatology, Allergy and Immunology, Chang Gung Memorial Hospital at Linkou, No. 5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan

5 Division of Rheumatology, Orthopaedics and Dermatology, School of Medicine, University of Nottingham, Nottingham NG7 2UH, UK

6 School of Medicine, Chang Gung University, No. 259, Wenhua 1st Rd., Guishan Dist., Taoyuan City 33302, Taiwan

7 Department of Education, Chang Gung Memorial Hospital at Linkou, No.5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan

8 Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital, Linkou, No. 5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan

9 College of Medicine, Institute for Radiologic Research, Chang Gung University, No. 259, Wenhua 1st Rd., Guishan Dist., Taoyuan City 33302, Taiwan

10 Center of Vascularized Tissue Allograft, Gung Memorial Hospital at Linkou, No. 5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan

11 Healthy Aging Research Center, Chang Gung University, No. 259, Wenhua 1st Rd., Guishan Dist., Taoyuan City 33302, Taiwan