



OPEN ACCESS

EDITED BY

Eduardo Abreu,
University of Missouri–Kansas City,
United States

REVIEWED BY

Roland Celoud,
State Hospital Horn, Austria
Fusong Jiang,
Shanghai Jiao Tong University, China
Kamil Koszela,
National Institute of Geriatrics, Rheumatology
and Rehabilitation, Poland

*CORRESPONDENCE

Satoshi Kasai
✉ satoshi-kasai@nuhw.ac.jp

RECEIVED 21 September 2024

ACCEPTED 07 March 2025

PUBLISHED 26 March 2025

CITATION

Yoshida A, Sato Y, Kai C, Hirono Y, Sato I and
Kasai S (2025) Utility of osteoporosis
screening based on estimation of bone
mineral density using bidirectional chest
radiographs with deep learning models.
Front. Med. 12:1499670.
doi: 10.3389/fmed.2025.1499670

COPYRIGHT

© 2025 Yoshida, Sato, Kai, Hirono, Sato and
Kasai. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Utility of osteoporosis screening based on estimation of bone mineral density using bidirectional chest radiographs with deep learning models

Akifumi Yoshida¹, Yoichi Sato², Chiharu Kai^{1,3}, Yuta Hirono^{3,4},
Ikumi Sato^{3,5} and Satoshi Kasai^{1*}

¹Department of Radiological Technology, Faculty of Medical Technology, Niigata University of Health and Welfare, Niigata, Japan, ²Nagoya University Graduate School of Medicine, Aichi, Japan, ³Major in Health and Welfare, Graduate School of Niigata University of Health and Welfare, Niigata, Japan, ⁴TOITU Co., Ltd., Tokyo, Japan, ⁵Department of Nursing, Faculty of Nursing, Niigata University of Health and Welfare, Niigata, Japan

Introduction: Osteoporosis increases the risk of fragility fractures, especially of the lumbar spine and femur. As fractures affect life expectancy, it is crucial to detect the early stages of osteoporosis. Dual X-ray absorptiometry (DXA) is the gold standard for bone mineral density (BMD) measurement and the diagnosis of osteoporosis; however, its low screening usage is problematic. The accurate estimation of BMD using chest radiographs (CXR) could expand screening opportunities. This study aimed to indicate the clinical utility of osteoporosis screening using deep-learning-based estimation of BMD using bidirectional CXRs.

Methods: This study included 1,624 patients aged ≥ 20 years who underwent DXA and bidirectional (frontal and lateral) chest radiography at a medical facility. A dataset was created using BMD and bidirectional CXR images. Inception-ResNet-V2-based models were trained using three CXR input types (frontal, lateral, and bidirectional). We compared and evaluated the BMD estimation performances of the models with different input information.

Results: In the comparison of models, the model with bidirectional CXR showed the highest accuracy. The correlation coefficients between the model estimates and DXA measurements were 0.766 and 0.683 for the lumbar spine and femoral BMD, respectively. Osteoporosis detection based on bidirectional CXR showed higher sensitivity and specificity than the models with single-view CXR input, especially for osteoporosis based on T-score ≤ -2.5 , with 92.8% sensitivity at 50.0% specificity.

Discussion: These results suggest that bidirectional CXR contributes to improved accuracy of BMD estimation and osteoporosis screening compared with single-view CXR. This study proposes a new approach for early detection of osteoporosis using a deep learning model with frontal and lateral CXR inputs. BMD estimation using bidirectional CXR showed improved detection

performance for low bone mass and osteoporosis, and has the potential to be used as a clinical decision criterion. The proposed method shows potential for more appropriate screening decisions, suggesting its usefulness in clinical practice.

KEYWORDS

bone mineral density, osteoporosis, screening, chest radiograph, artificial intelligence

1 Introduction

Osteoporosis is a systemic bone disease that causes bone fragility due to loss of bone mass (1). Osteoporosis is the most significant risk factor for fragility fractures and has a substantial impact on life prognosis (1–3). Among these, lumbar spine and femur fractures significantly worsen quality of life (4). In addition, osteoporosis typically does not cause symptoms until a fracture occurs (5, 6). Delayed diagnosis of osteoporosis leads to fragility fractures, with an estimated 37 million fragility fractures occurring annually in people aged 55 years and older worldwide between 1990 and 2019 (7). Therefore, early detection of osteoporosis before its progression is essential to prevent fragility fractures (8, 9).

Cost-effective screening is required to reduce the incidence of fragility fractures caused by osteoporosis (10). Osteoporosis is examined by measuring bone mineral density (BMD). Dual-energy X-ray absorptiometry (DXA) is the most accurate method for measuring BMD and remains the gold standard for screening osteoporosis (11, 12). Individual BMD measurements are assessed based on differences and ratios against the young adult mean (YAM) in skeletally healthy young adults. The T-score is an assessment index widely used as an international standard, and is the difference in an individual's BMD measurement against the YAM divided by the standard deviation of young adults' BMD. Corresponding to the World Health Organization (WHO) diagnostic categories, $-2.5 \geq T\text{-score}$ indicates osteoporosis, and $-1 > T\text{-score} > -2.5$ indicates low bone mass (13). In some countries, the relative ratio of an individual's BMD to YAM is also used; for example, the diagnostic criteria for osteoporosis and osteopenia in Japan define osteoporosis as a BMD less than 70% of the YAM (14). Thus, osteoporosis can be objectively diagnosed based on BMD measurements. However, the high cost of DXA equipment limits the availability of medical facilities that can perform DXA, and the low uptake of DXA measurements poses a challenge for the early detection of osteoporosis (13).

Effective screening opportunities should be expanded for the early detection of osteoporosis. It would be useful if BMD could be obtained from popular medical data other than bone mineral density testing, the risk of osteoporosis could be assessed, or medical examinations could be encouraged based on objective data. Previous studies have shown that the values obtained by analyzing anatomical features, such as the cortical thickness of the clavicle, spine, and ribs on radiographs, correlate with BMD and bone mass (15–17). Furthermore, many recent studies have used deep learning and other artificial intelligence (AI) methods to analyze medical images and obtain information for purposes other than the original

use. Previous studies have used deep learning models to estimate BMD using radiographic images of the lumbar spine and hip joint as inputs (18, 19) and to identify the presence of osteoporosis (20, 21), indicating the possibility of identifying osteoporosis with high accuracy. However, because lumbar and hip radiography are performed mainly as adjuncts to orthopedic consultations, the target population for the method of osteoporosis screening using lumbar and hip radiographs is limited to orthopedic patients. It has been reported that $\geq 75\%$ of all individuals for whom DXA testing is recommended do not undergo it (22–27). Thus, it remains important to promote screening more broadly to target individuals without orthopedic consultation or those who have not undergone DXA testing. A method that utilizes medical data obtained at a high frequency regardless of orthopedic symptoms could effectively extend osteoporosis screening opportunities.

Chest radiography is performed consistently in primary care and during health checkups and is the most widely and frequently performed basic diagnostic imaging technique worldwide. It would be promising to expand screening without requiring additional imaging if chest radiographs (CXRs) could be used to accurately

TABLE 1 Summary of data characteristics.

Demographics		
Sex, <i>n</i> (%)		
Male	2,682	(41.6%)
Female	3,764	(58.4%)
Age, mean [IQR] (years)	60.0	[47, 75]
BMI, mean [IQR] (kg/m ²)	20.3	[17.7, 22.2]
Bone mineral density		
Lumbar BMD, mean [IQR] (g/cm ²)	0.893	[0.770, 1.012]
Femoral BMD, mean [IQR] (g/cm ²)	0.616	[0.516, 0.711]
T-score, <i>n</i> (%)		
T-score ≥ -1.0	1,373	(21.3%)
$-2.5 < T\text{-score} < -1.0$	2,411	(37.4%)
T-score ≤ -2.5	2,662	(41.3%)
BMD/YAM, <i>n</i> (%)		
BMD/YAM $\geq 80\%$	2,650	(41.1%)
$70\% < \text{BMD/YAM} < 80\%$	1,373	(21.3%)
BMD/YAM $\leq 70\%$	2,423	(37.6%)

Age, BMI, lumbar BMD, and femoral BMD are shown as mean [25, 75 percentiles]. *n* is the number of applicable data-pairs and not the number of participants. BMI, body mass index; BMD, bone mineral density; IQR, Interquartile range; YAM, young adult mean.

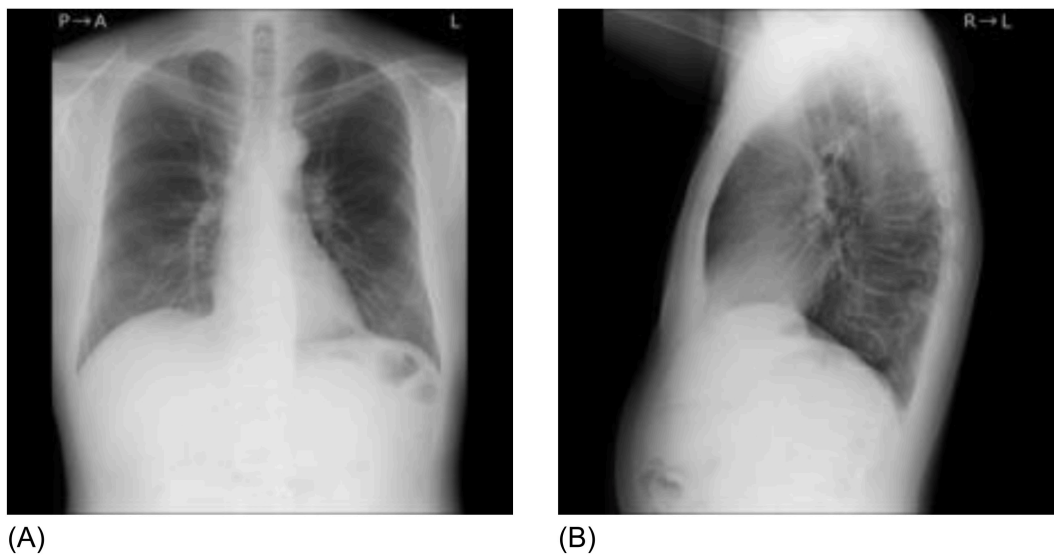


FIGURE 1

Sample chest radiographs of one patient used in this study. (A) Frontal chest radiograph, (B) Lateral chest radiograph. The original images were down-sampled and zero-padded to $1,024 \times 1,024$ matrixes with the aspect ratio preserved and resized to 299×299 as the input size for the pre-trained model.

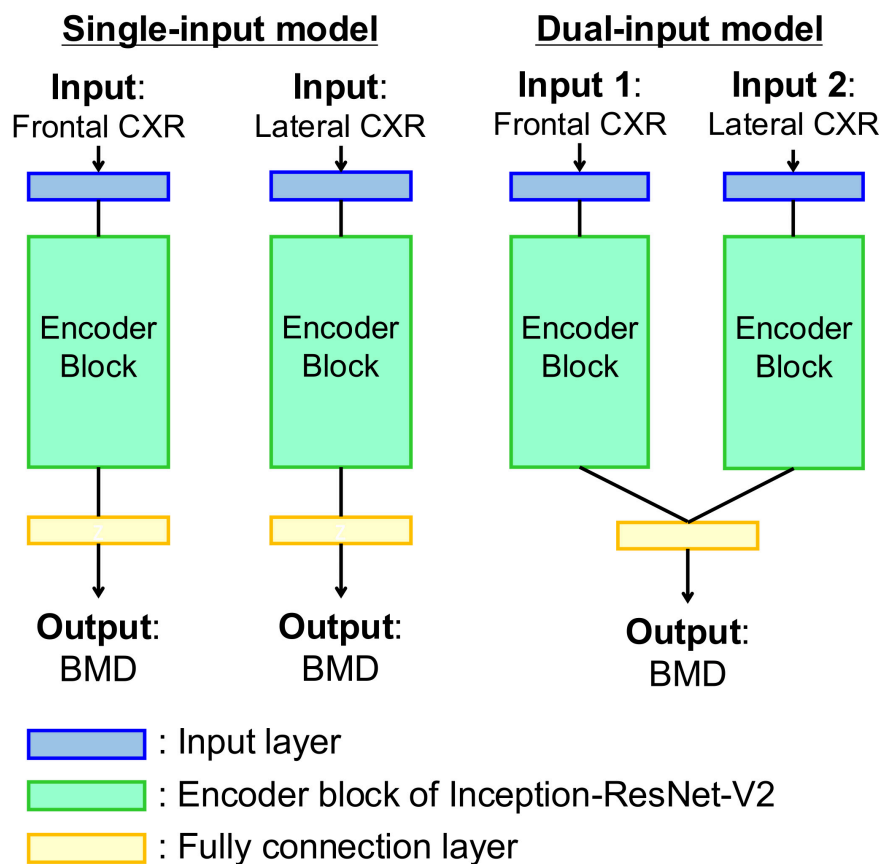


FIGURE 2

Diagrams of the model for estimating bone density from chest radiographs. (Left) Single-input model with input of frontal or lateral chest radiograph. (Right) Dual-input model with inputs of bidirectional chest radiographs. Each encoder block is composed of a pre-trained Inception-ResNet-V2 network as feature extractor.

assess the risk (28). Studies investigating approaches to detect osteoporosis from CXR have been reported, including studies using deep learning models that directly classified the prevalence of osteoporosis from CXR (29) and estimated BMD from CXR with patient information (30, 31). This BMD estimation approach would be more valuable for flexible use in clinical practice because BMD estimates could be used to identify osteoporosis and/or to consider the necessity for DXA testing. Sato et al. reported the detection of osteopenia with a cutoff T-score of -1.0 , based on estimated BMD, and obtained a sensitivity of 90.14% (30). Meanwhile, the detection sensitivity for osteoporosis with a cutoff T-score of -2.5 was relatively low at 77.27%. A study evaluating the recommendation criteria for osteoporosis screening described a sensitivity of $> 90\%$ as necessary (32); for example, the National Osteoporosis Foundation guideline criteria in the United States were evaluated and found to yield a sensitivity of 96.2% and specificity of 17.8% for selecting cases with T-score ≤ -2.5 in postmenopausal women aged ≥ 45 years (32). Improving the sensitivity of osteoporosis detection is essential to promote adequate osteoporosis screening opportunities and to provide more persuasive recommendations for medical examinations in recipients undergoing chest radiography.

In primary care and health checkups in many countries, chest radiography with frontal and lateral bidirectional imaging is often performed depending on the patient's symptoms (33). A lateral CXR can visualize the thoracic region up to the upper lumbar spine, and the shape of the vertebrae can be observed. Imaging findings of fragility fractures such as vertebral compression fractures are also diagnostic criteria for osteoporosis (13), and lateral CXRs are effective for detecting vertebral fractures (34). Previous studies using lateral CXR have been reported on osteoporosis screening using lateral CXR with automatic detection using AI (35). Previous studies using AI to estimate BMD from medical images tended to show higher estimation performance when using radiographs in which the region to be measured by DXA was imaged than when it was not (36). Therefore, we hypothesized that AI with input of lateral CXR can estimate BMD effectively, and that combining frontal and lateral CXR images will improve the accuracy of AI estimation compared with using only the frontal image as input.

This study aimed to develop deep learning models to estimate BMD using only bidirectional CXR as input and to clarify their performance improvement over models using only frontal CXR as input, as well as their utility in osteoporosis screening.

2 Materials and methods

2.1 Materials

2.1.1 Data

This study was approved by the institutional review boards of the participating institutions (approval number: 18973-230111). The Institutional Review Board determined that formal informed consent was not required because this study used de-identified clinical data.

TABLE 2 Prediction results for bone mineral density of the lumbar spine and femoral neck by deep learning models with three different input type.

	R [95% CI]		MAPE [95% CI]/%	
Lumbar BMD				
Frontal	0.709	[0.682–0.707]	11.5	[11.3–11.8]
Lateral	0.732	[0.720–0.743]	11.0	[10.8–11.3]
Bidirectional	0.766	[0.756–0.776]	10.6	[10.4–10.9]
Femoral BMD				
Frontal	0.581	[0.565–0.597]	14.7	[14.3–15.1]
Lateral	0.624	[0.609–0.638]	14.7	[14.3–15.0]
Bidirectional	0.683	[0.670–0.696]	13.8	[13.5–14.1]

BMD, bone mineral density; CI, confidence interval; R, Pearson's correlation coefficient; MAPE, mean absolute percentage error. Bolded values of R and MAPE represent the best value for each BMD.

This study included patients aged ≥ 20 years who underwent bidirectional CXR and BMD measurement using DXA within 1 year at a single medical facility in Japan from April 2010 to July 2022. This study included DXA examinations performed at the facility during the study period and corresponding chest radiographs (frontal and lateral) performed at the facility within 1 year before and after the DXA examination. All the paired data were incorporated within the inclusion period. Data were excluded if the CXR lacked parts of the lungs or clavicles and if the CXR was performed using portable devices. A total of 1,624 cases (520 males, 1,104 females) with 6,446 data pairs (2,682 males, 3,764 females) of multiple BMD measurements and bidirectional CXR images met these criteria. The data attributes are listed in Table 1. A Horizon (Hologic Inc., Marlborough, MA, United States) was used for BMD measurements. Lumbar BMD measured in the lumbar vertebral region and femoral BMD measured in the femoral neck region were targeted. T-scores and BMD/YAM were calculated based on the mean BMD values and standard deviations for each sex of young adults, corresponding to the DXA device used for actual BMD measurements. According to the WHO (Geneva) criteria, the participants were classified based on their T-scores into normal (T-score ≥ -1.0), osteopenia ($-1.0 < \text{T-score} < -2.5$), and osteoporosis (T-score ≤ -2.5) groups (37). All bidirectional CXRs used in this study were obtained using frontal and lateral chest radiography. Figure 1 shows an example of a bidirectional CXR. Each original CXR image of different sizes was resized while preserving the aspect ratio and was zero-padded to a $1,024 \times 1,024$ matrix.

2.1.2 Dataset splitting

A 10-fold cross-validation method was used to create the datasets; each of the 10 datasets was created by randomly splitting the participants into training, validation, and evaluation ratios of 80%, 10%, and 10%, respectively. In each dataset used for cross validation, there was no contamination by the same participants between the training, validation, and evaluation data. The AI outputs for the ten evaluation datasets were combined, and all data were used for evaluation.

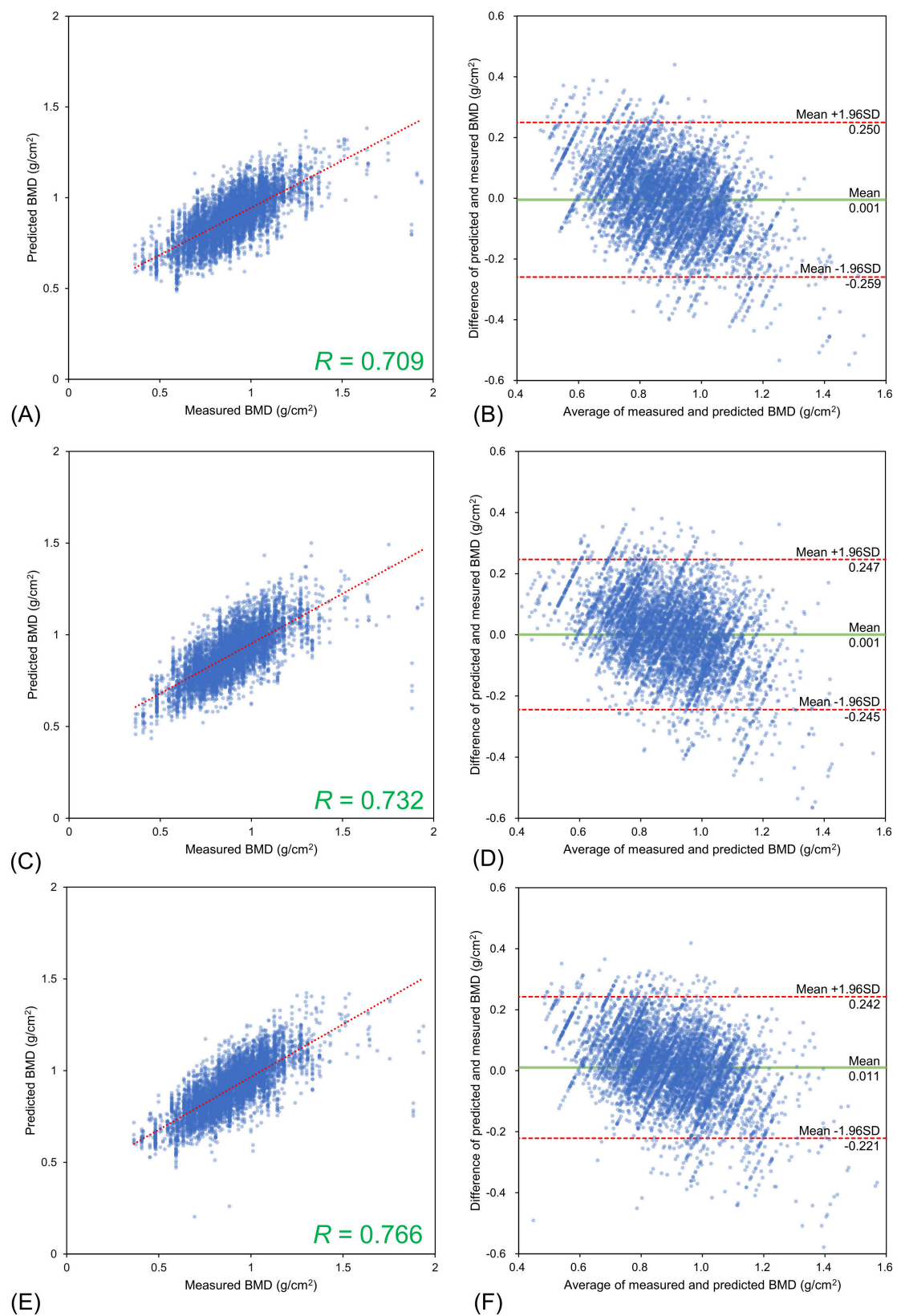


FIGURE 3

Estimation results for bone mineral density of the lumbar spine using three different input types. Upper (A,B) Frontal chest radiographs (CXR) input, Middle (C,D) Lateral CXR input, Bottom (E,F) Bidirectional CXR input. Left (A,C,E) Relationships of measured and estimated values, Right (B,D,F) Bland–Altman plots. All the models confirmed the agreement between model estimates and dual-energy absorptiometry measurements.

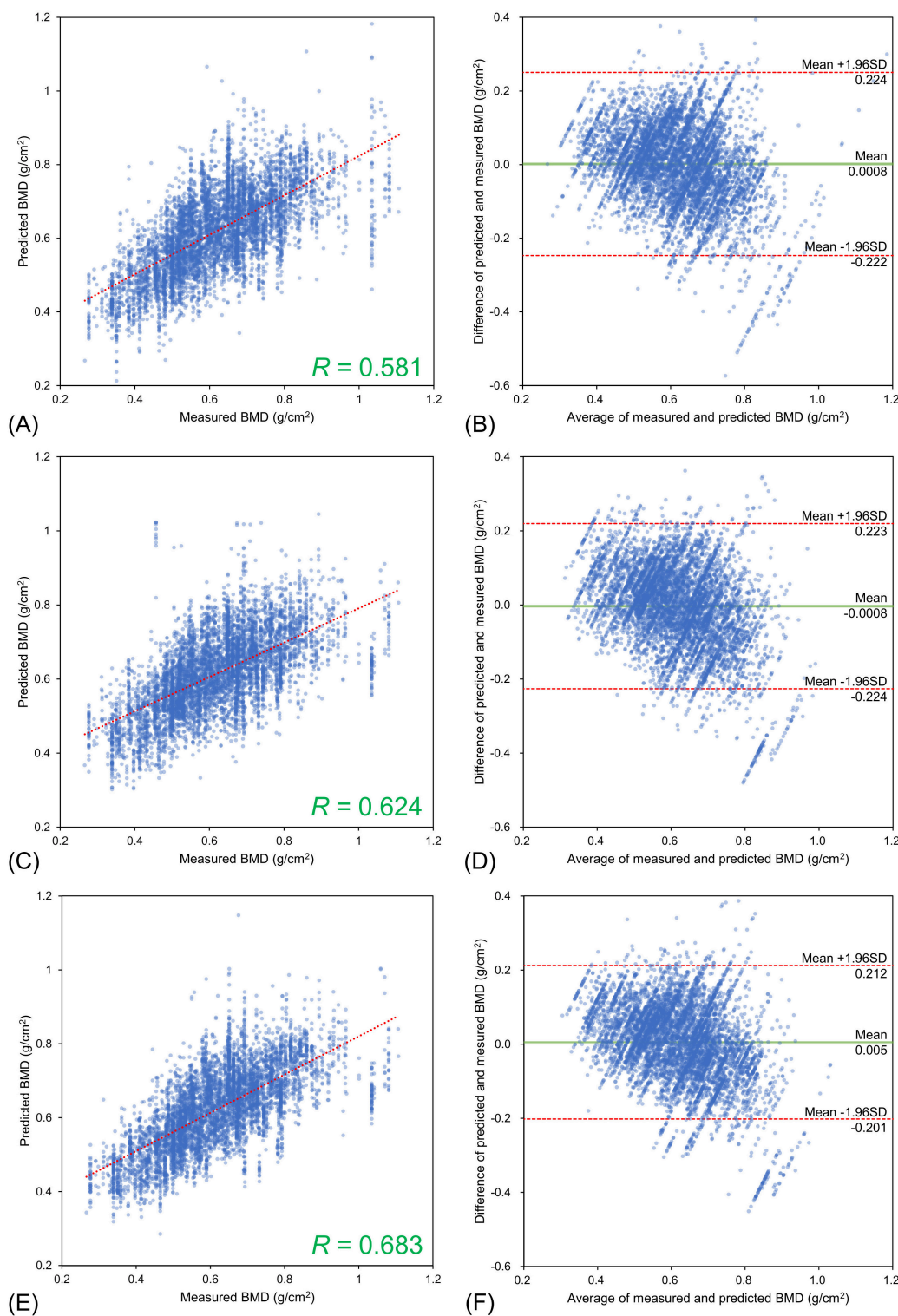


FIGURE 4
 Estimation results for bone mineral density of the femoral neck using three different input types. Upper (A,B) Frontal chest radiographs (CXR) input, Middle (C,D) Lateral CXR input, Bottom (E,F) Bidirectional CXR input. Left (A,C,E) Relationships of measured and estimated values, Right (B,D,F) Bland–Altman plots. All the models confirmed the agreement between model estimates and dual-energy absorptiometry measurements.

2.1.3 Experimental environment

The specifications of the experimental computer were an Intel Core i5-12400 CPU, 16 GB × 2 of RAM, and an NVIDIA GeForce RTX 3090 GPU with a VRAM of 24 GB.

The experiment was implemented using MATLAB 2023b (MathWorks, Inc.) on Windows 11 (Microsoft, Inc.). All image processing and deep learning network computations were performed using MATLAB.

TABLE 3 Summary of statistics in Bland–Altman analysis.

	Mean difference [95% CI] (g/cm ²)	Limits of agreement	
		Range (g/cm ²)	Agreement/%
Lumbar BMD			
Frontal	−0.0011	[−0.0079 to −0.0016]	−0.262 to 0.250 96.1
Lateral	0.0009	[−0.0021 to 0.0040]	−0.245 to 0.247 95.6
Bidirectional	0.0106	[0.0077 to 0.0135]	−0.221 to 0.242 95.5
Femoral BMD			
Frontal	0.0017	[−0.0014 to 0.0048]	−0.247 to 0.250 97.2
Lateral	−0.0033	[−0.0061 to −0.0005]	−0.226 to 0.220 95.6
Bidirectional	0.0052	[0.0026 to 0.0078]	−0.202 to 0.212 95.7

CI, confidence interval; BMD, bone mineral density.

TABLE 4 Detection results with predicted values by different models for osteoporosis and osteopenia based on T-score of measured bone mineral density.

	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC/%
BMD T-score < −1.0						
Frontal	79.5	89.1	43.8	85.4	52.1	79.4%
Lateral	78.9	90.3	36.9	84.1	50.7	79.0%
Bidirectional	80.6	90.0	46.0	86.0	55.4	81.5%
BMD T-score ≤ −2.5						
Frontal	74.5	64.7	81.4	71.0	76.6	82.3%
Lateral	72.6	62.1	79.9	68.5	75.0	80.7%
Bidirectional	76.5	63.4	85.6	75.6	76.9	84.2%

Accuracy, Sensitivity, Specificity, PPV, NPV, and AUROC are shown as percentage values. BMD, bone mineral density; PPV, positive predictive value; NPV, negative predictive value; AUROC, area under the receiver operating characteristic curve. Bolded values represent the best in each BMD T-score category.

2.2 Methods

2.2.1 Experiments

The Inception-ResNet-V2 model (38) was used as the feature extractor. The model data were acquired from the public MATLAB add-on library (39). The model data were used as a pretrained network obtained from a classification task using ImageNet (40). The classifier in the base model was replaced by a single fully connected layer, with a single output class as the regression layer. The models used in the experiments were a single-input model, which used one direction of the CXR as the input, and a dual-input model, which used a bidirectional CXR as the input. Figure 2 shows a structural diagram of the models. Using the single-input model, training was conducted to estimate the BMD using the CXR frontal or lateral views as the input. The dual-input model was implemented using two single-input models trained on frontal and lateral CXR. The trained feature extractors of the single-input models were connected in parallel to the fully connected layer. The dual-input model was trained to estimate the BMD using paired frontal and lateral CXRs inputs. The training conditions were as follows: optimization method Adam, loss function root mean square error, learning rate 1×10^{-3} , 3×10^{-4} , 1×10^{-4} , 3×10^{-5} , 1×10^{-5} (variable), batch size 32, maximum number of epochs 100, and image data augmentation with $\pm 5^\circ$ rotation, horizontal flipping, and $\pm 5\%$ scaling. The learning rate was selected to obtain the lowest loss-function value for each model. The input images

were resized to the input size of the pretrained model (299×299 pixels) during image data augmentation and then used as the model input. The training dataset was used to update the network weights and the validation dataset was used to display the model performance at each epoch. Training was stopped early when the loss value for the validation dataset at the end of each epoch did not update the minimum for 10 consecutive epochs. The weights at the epoch with the smallest loss function output on the validation dataset were saved, thereby completing the training. Model training was conducted separately for each target to be estimated as the lumbar and femoral BMD.

2.2.2 Evaluation

The trained model was used to input chest X-ray images from the evaluation dataset and output BMD estimates. The Pearson's correlation coefficient (R) and mean absolute percentage error (MAPE) between the reference measured values and model estimates were used as the evaluation metrics. A 95% confidence interval (CI) of the correlation coefficient was obtained using Fisher's z-transform. A Bland–Altman analysis was performed using the mean values of the reference measured values and model estimates and the differences between the estimates and measured values. If the following conditions were met, the estimated values and measured values were considered equivalent: (1) The 95% CI of the mean difference between the estimated and measured values included ± 0.01 g/cm², and (2) more than 95% of the evaluation

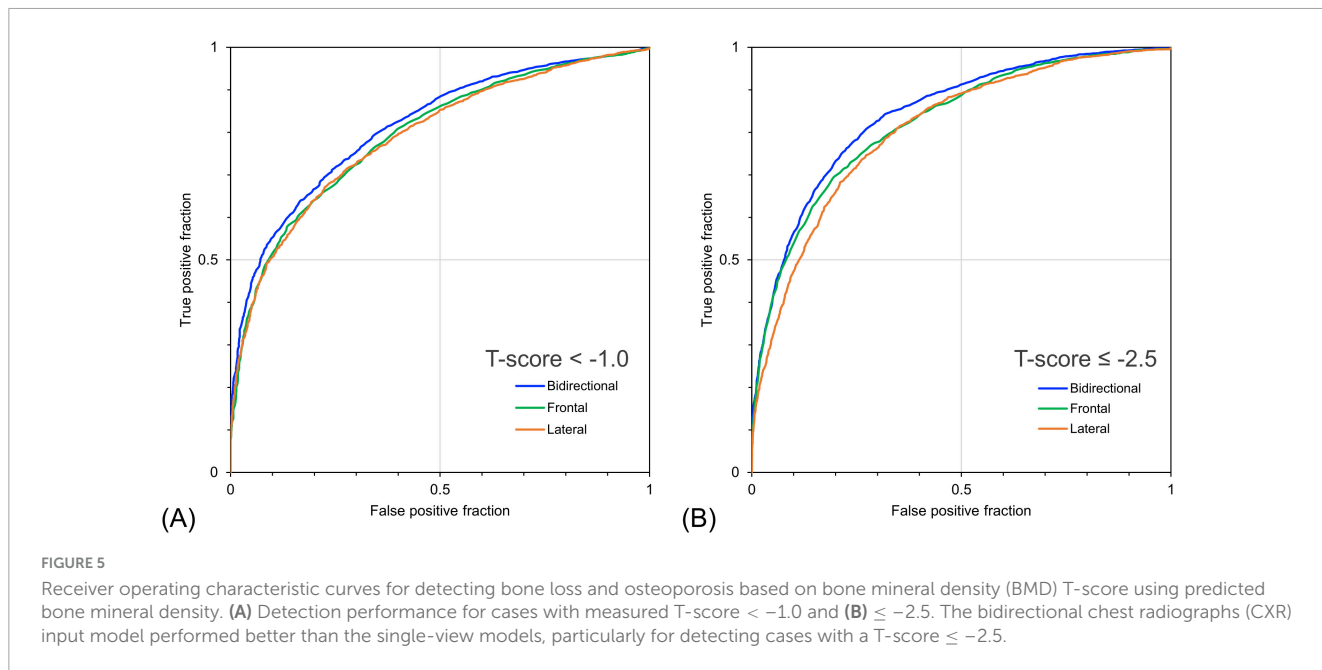


TABLE 5 Detection results with predicted values by different models for osteoporosis and osteopenia based on the ratio of measured bone mineral density to the young adult mean.

	Accuracy	Sensitivity	Specificity	PPV	NPV	AUROC/%
BMD/YAM < 80%						
Frontal	69.4	75.3	61.0	73.5	63.3	77.2%
Lateral	71.0	79.1	59.5	73.6	66.5	78.0%
Bidirectional	71.6	75.2	66.3	76.2	65.1	79.7%
BMD/YAM $\leq 70\%$						
Frontal	76.2	61.5	85.0	71.2	78.6	83.4%
Lateral	74.9	58.7	84.6	69.6	77.3	82.0%
Bidirectional	77.8	60.8	88.0	75.4	78.9	85.6%

Accuracy, Sensitivity, Specificity, PPV, NPV, and AUROC are shown as percentage values. PPV, positive predictive value; NPV, negative predictive value; AUROC, area under the receiver operating characteristic curve; BMD, bone mineral density; YAM, the young adult mean. Bolded values represent the best in each BMD/YAM category.

data were included within the limits of agreement (LOA) of mean ± 1.96 SD (41).

The predicted T-score and BMD/YAM ratio were calculated using the predicted BMD and actual YAM according to sex. Based on the WHO guidelines, osteoporosis (measured T-score ≤ -2.5) and osteopenia (measured T-score < -1.0) were detected from the calculated predicted T-scores. Additionally, the detection performance for osteoporosis and osteopenia was evaluated. Furthermore, based on Japanese guidelines (13), measured BMD/YAM $\leq 70\%$ (osteoporosis) and $< 80\%$ (low bone mass) were detected from the predicted BMD/YAM. The evaluation metrics included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic curve (AUROC).

3 Results

The results of the BMD estimation for each model with different inputs are presented in Table 2. The model with

bidirectional CXR inputs showed superior correlation coefficients to models with only frontal or lateral image inputs: $R = 0.766$ (95% CI 0.756–0.776) and 0.683 (95% CI 0.670–0.696) for lumbar and femoral BMD estimation, respectively.

The lumbar BMD estimation results for each input image type relative to the DXA measurements are shown in Figure 3 and the femoral BMD estimation results are shown in Figure 4. The statistics for the Bland–Altman analysis are shown in Table 3. In the Bland–Altman analysis based on lumbar BMD estimates by the three models and actual measurements by DXA, the 95% CIs for the mean of the differences between the estimates and measurements overlapped with the reference range, indicating no fixed errors. The percentage of data within the LOA was greater than 95% for all three models. These results confirm the agreement between the lumbar BMD estimates and measured values in the three models with different input types. Bland–Altman analysis for femoral BMD estimation showed that the 95% CIs of the mean differences between the model estimates and the DXA measurements overlapped with the reference range, indicating no fixed errors. The percentage of data within the LOA was greater

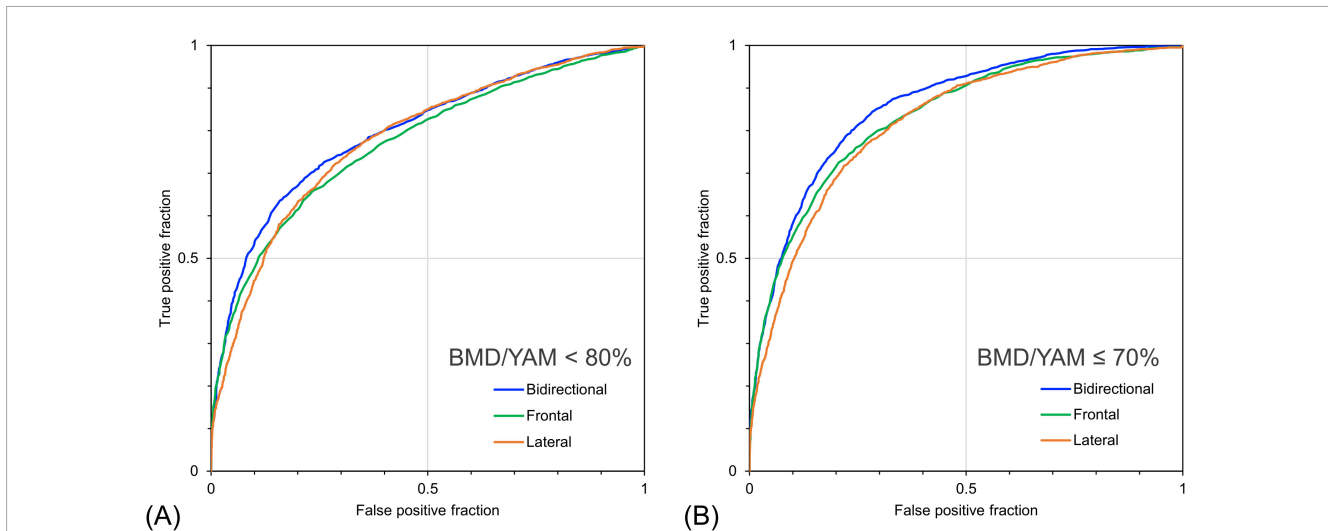


FIGURE 6

Receiver operating characteristic curves for detecting bone loss and osteoporosis based on the ratio of individual bone mineral density to the young adult mean (BMD/YAM) using predicted bone mineral density. (A) Detection performance for cases with measured BMD/YAM < 80% and (B) < 70%. The bidirectional chest radiographs (CXR) input model performed better than the single-view models, particularly for detecting cases with BMD/YAM <= 70%.

TABLE 6 Sensitivities of the models with different inputs in detecting risk groups for osteoporosis based on measured BMD, with variable cutoffs based on model estimates and tuned specificity.

	Sensitivity/%	
	With 40% specificity	With 50% specificity
BMD T-score <math>\leq -2.5</math>		
Frontal	93.5	88.8
Lateral	92.4	89.1
Bidirectional	94.5	91.3
BMD T-score <math>< -1.0</math>		
Frontal	90.1	86.1
Lateral	89.7	85.3
Bidirectional	92.0	88.4
BMD/YAM <math>\leq 70\%</math>		
Frontal	94.9	90.7
Lateral	93.7	91.0
Bidirectional	95.8	92.8

Bolded values represent sensitivity above 90%, Underlined values are the highest values for the three input types. BMD, bone mineral density; YAM, the young adult mean.

than 95% for all three models. These results confirm the agreement between the femoral BMD estimates and actual measurements for the three models with different input types.

Table 4 shows the detection performance for low bone mass and osteoporosis using the estimated T-score obtained by the models with different inputs. The model with the bidirectional CXR input yielded the highest values for all other indicators, except for sensitivity. In the model with bidirectional CXR inputs, the detection performance for T-score <math>< -1.0</math> and T-score ≤ -2.5 were 90.0% and 63.4% for sensitivity, 46.0% and 85.6% for specificity, and 81.5% and 84.2% for AUROC, respectively. The

model with bidirectional CXR inputs improved specificity and AUROC by approximately 2.2% and 2.1% for T-score <math>< -1.0</math> and 4.2% and 1.9% for T-score ≤ -2.5, respectively, compared with the model with frontal CXR input. Figure 5 shows the ROC curves for detecting low bone mass (measured T-score <math>< -1.0</math>) and osteoporosis (measured T-score ≤ -2.5) using predicted T-scores. For osteoporosis detection, the model with bidirectional CXRs showed better performance than those with only frontal or lateral CXRs.

Table 5 shows the detection performance for low bone mass and osteoporosis based on the estimated BMD/YAM derived from the models with different inputs. The model with bidirectional CXR yielded the highest values for almost all indices except for sensitivity. The model with bidirectional CXR improved the specificity and AUROC for detecting osteopenia with a BMD/YAM cutoff of 80% by 5.3% and 2.5%, respectively, and for detecting osteoporosis with a BMD/YAM cutoff of 70% by 3.0% and 2.2%, respectively, compared with the model with only frontal CXR input. Figure 6 shows the ROC curves for the detection of low bone mass and osteoporosis using the predicted BMD/YAM ratios. The model with bidirectional CXR showed higher performance for osteoporosis detection than those with only frontal or lateral images.

4 Discussion

The results of this study suggest that ensemble learning models with frontal and lateral CXR inputs yield a higher BMD estimation accuracy than models with single-view CXR inputs. To the best of our knowledge, this is the first study to develop a deep learning model to estimate BMD using only lateral and frontal CXR images. The results of the ensemble model with bidirectional CXR inputs indicated higher accuracy of osteoporosis screening in clinical practice. Furthermore, our findings add value to the previous

studies on BMD estimation using chest radiographs as an input. The lateral CXR images contained information related to BMD, and the effectiveness of the lateral CXR images for estimating BMD using AI was demonstrated.

The results of lumbar spine BMD estimation using bidirectional CXR (frontal and lateral) showed a strong positive correlation of $R = 0.766$ with the actual DXA measurements, indicating the feasibility of obtaining a stronger correlation than that obtained in previous studies. Among previous studies on BMD estimation using chest radiographs, Sato et al. reported the highest estimation performance (30). Using CXR frontal images and patient information (age and sex) as model inputs, they obtained $R = 0.68$ for lumbar BMD estimation and $R = 0.75$ for femoral BMD estimation. Meanwhile, the estimation of femoral BMD using bidirectional CXRs in this study showed a moderate correlation of $R = 0.683$, which was relatively weak. The average of 5,157 training data points in the 10-fold cross-validation in this study was about 58.8% less than the 12,529 training data points used by Sato et al. Generally, the performance of deep learning models increases with the number of training data points. Our model with bidirectional CXR inputs has the potential to achieve even higher estimation performance with greater amounts of training data.

The developed dual-input model with bidirectional CXR inputs is more effective than the single-input model in terms of application to osteoporosis screening. The estimates from the model with bidirectional CXR inputs in this study identified the presence of a measured T-score of < -1.0 , with 90% sensitivity and 46% specificity, and a cutoff predicted T-score of -1.0 . Regarding triage screening for osteoporosis, a sensitivity of $> 90\%$ and a specificity of approximately 40%–60% or higher are considered acceptable clinical decision criteria (32), and our results with a fixed cutoff fulfilled these criteria for the identification of a T-score < -1.0 . Furthermore, suppose that the cut-off of the predicted T-score based on the model output is tuned according to this criterion. In such cases, it can provide even higher sensitivity for the detection of osteoporosis. For the detection task for cases with T-score ≤ -2.5 , the cutoff of predicted BMD was modified to 40% and 50% specificity. Table 6 lists the sensitivities of the models with different input information when the specificity was tuned. The model with bidirectional CXR inputs showed a sensitivity of 94.5% at a specificity of 40% and a sensitivity of 91.3% at a specificity of 50% for detecting a measured T-score ≤ -2.5 . Thus, the developed model may be helpful for screening for osteoporosis and low bone mass. The sensitivity of the model with bidirectional CXR as input at 50% specificity was 91.3% for T-score ≤ -2.5 , which was + 2.5% higher than that of the model with frontal CXR as input. The model with only frontal CXR showed a sensitivity of 88.8%, which was less than acceptable for the clinical decision criteria. These results indicate the usefulness of deep learning methods using bidirectional CXR for osteoporosis screening.

The results of the estimates from the bidirectional CXR model showed the feasibility of effective screening for osteoporosis and osteopenia based on the T-score and BMD/YAM criteria. A T-score ≤ -2.5 is a current global diagnostic criterion for osteoporosis. However, other criteria have been used to diagnose osteoporosis in other regions. For example, the Japanese criteria are based on BMD/YAM (37). Future diagnostic criteria may change based on the medical conditions. In addition, the management

of osteoporosis risk groups is essential. Therefore, it is important to provide a robust detection performance for diagnoses based on various cutoff values. As shown in Table 6, the results of this study with bidirectional CXRs met the sensitivity $> 90\%$ and specificity $> 40\%$ levels for multiple cutoffs of T-score ≤ -2.5 , T-score < -1.0 , and BMD/YAM $\leq 70\%$. Thus, BMD prediction using bidirectional CXRs can provide robust screening performance for osteoporosis and its risk groups based on multiple cut-off values.

This study used only CXR images as model input to estimate BMD and did not use patient information such as age, sex, or clinical covariates. Previous studies estimating BMD from CXR used a model with frontal CXR and patient information (30, 31), although it was not clarified to what extent the AI model performance was affected by the input of patient information. The reason patient information was not input in this study was to eliminate the influence of the input of patient information on the estimation results since the purpose was to clarify the differences in performance depending on the input of CXR imaging direction and number of inputs. Another study reported that the use of patient information such as age, sex, and body mass index improved model performance (42). In future studies, our model may improve the accuracy of osteoporosis screening in primary care and health check-ups by inputting patient information and clinical covariates. This study aimed to propose an expansion of opportunity for osteoporosis screening by estimating BMD from chest radiographs, which are frequently obtained in primary care. Regarding opportunistic screening using chest radiographs, studies have reported on the estimation of pulmonary function (43–45) and the prediction of cardiovascular events (46) using deep learning models. These studies have proposed the secondary use of chest radiographs acquired for other purposes to incidentally detect the risk of lifestyle-related diseases. This study could be applied concurrently with these approaches to provide an opportunity to intervene before disease progression.

This study had two limitations. Firstly, this study used data obtained from a single facility to train and evaluate the model, which may have been affected by the regional characteristics. To validate the generalizability of this study's findings, it is important to evaluate their performance using data collected from different facilities and/or racial groups. Secondly, the diagnostic results of chest radiography were not considered. The results of this study could potentially have been influenced by imaging findings; however, their influence is unclear.

5 Conclusion

This study developed a deep learning model to estimate BMD using frontal and lateral CXRs and demonstrated the utility of the model for osteoporosis screening based on the model estimates. The model with bidirectional CXR inputs showed higher BMD estimation performance than the model with a single CXR input. This suggests the usefulness of a BMD estimation model using bidirectional CXR inputs for screening for osteoporosis and low bone mass.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of Niigata University of Health and Welfare. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because this study constitutes an observational study without individually identifiable information.

Author contributions

AY: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. YS: Conceptualization, Data curation, Resources, Supervision, Writing – review and editing. CK: Resources, Validation, Writing – review and editing. YH: Validation, Writing – review and editing. IS: Validation, Writing – review and editing. SK: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review and editing.

References

- Compston J, McClung M, Leslie W. Osteoporosis. *Lancet*. (2019) 393:364–76. doi: 10.1016/S0140-6736(18)32112-3
- Ensrud K, Thompson D, Cauley J, Nevitt M, Kado D, Hochberg M, et al. Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass. Fracture intervention trial research group. *J Am Geriatr Soc*. (2000) 48:241–9. doi: 10.1111/j.1532-5415.2000.tb02641.x.3
- Nguyen N, Center J, Eisman J, Nguyen T. Bone loss, weight loss, and weight fluctuation predict the mortality risk in elderly men and women. *J Bone Miner Res*. (2007) 22:1147–54. doi: 10.1359/jbmr.070412
- John TS. Epidemiology of vertebral fractures. *J Clin Densitometry*. (2016) 19:8–22. doi: 10.1016/j.jocd.2015.08.004
- Sözen T, Özüşik L, Başaran NÇ. An overview and management of osteoporosis. *Eur J Rheumatol*. (2017) 4:46–56. doi: 10.5152/eurjrheum.2016.048
- Johnston C, Dagar M. Osteoporosis in older adults. *Med Clin North Am*. (2020) 104:873–84. doi: 10.1016/j.mcna.2020.06.004
- GBD 2019 Fracture Collaborators. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: A systematic analysis from the global burden of disease study 2019. *Lancet Healthy Longev*. (2021) 2:e580–92. doi: 10.1016/S2666-7568(21)00172-0
- Odén A, McCloskey E, Johansson H, Kanis J. Assessing the impact of osteoporosis on the burden of hip fractures. *Calcif Tissue Int*. (2013) 92:42–9. doi: 10.1007/s00223-012-9666-6
- Mitchell P. Fracture liaison services: The UK experience. *Osteoporos Int*. (2011) 22:487–94. doi: 10.1007/s00198-011-1702-2
- Rachner T, Khosla S, Hofbauer L. Osteoporosis: Now and the future. *Lancet*. (2011) 377:1276–87. doi: 10.1016/S0140-6736(10)62349-5
- US Preventive Services Task Force. Screening for osteoporosis to prevent fractures: US preventive services task force recommendation statement. *JAMA*. (2025) 6:498–508. doi: 10.1001/jama.2024.27154
- Kahwati L, Kistler C, Booth G, Sathe N, Gordon R, Okah E, et al. Screening for osteoporosis to prevent fractures: A systematic evidence review for the US preventive services task force. *JAMA*. (2025) 6:509–31. doi: 10.1001/jama.2024.21653.13
- World Health Organization. *WHO Criteria for Diagnosis of Osteoporosis*. 4BoneHealth. Geneva: WHO (2023).
- Soen S, Fukunaga M, Sugimoto T, Sone T, Fujiwara S, Endo N, et al. Diagnostic criteria for primary osteoporosis: Year 2012 revision. *J Bone Miner Metab*. (2013) 31:247–57. doi: 10.1007/s00774-013-0447-8
- Kumar D, Anburajan M. The role of hip and chest radiographs in osteoporotic evaluation among the south Indian women population: A comparative scenario with DXA. *J Endocrinol Invest*. (2014) 37:429–40. doi: 10.1007/s40618-014-0074-9
- Chen H, Zhou X, Fujita H, Onozuka M, Kubo K. Age-related changes in trabecular and cortical bone microstructure. *Int J Endocrinol*. (2013) 2013:213234. doi: 10.1155/2013/213234
- Holcombe S, Hwang E, Derstine B, Wang S. Measurement of rib cortical bone thickness and cross-section using CT. *Med. Image Anal Elsevier B.V.* (2018) 49:27–34.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (C) (Grant Number 24K146880), the Japanese Orthopedic Association (Grant Number 2022-3), and Grant-in-Aid for Research Formulation from the Niigata University of Health and Welfare (2024).

Acknowledgments

We would like to thank Editage (www.editage.jp) for language editing.

Conflict of interest

YS was previously the president of iSurgery Co., Ltd. YH was employed by TOITU Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

18. Yamamoto N, Sukegawa S, Kitamura A, Goto R, Noda T, Nakano K, et al. Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates. *Biomolecules*. (2020) 10:1534. doi: 10.3390/biom10111534
19. Zhang B, Yu K, Ning Z, Wang K, Dong Y, Liu X, et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study. *Bone*. (2020) 140:115561. doi: 10.1016/j.bone.2020.115561
20. Nguyen T, Chae D, Park S, Yoon JA. Novel approach for evaluating bone mineral density of hips based on a sobel gradient-based map of radiographs utilizing a convolutional neural network. *Comput. Biol. Med. Elsevier Ltd*. (2021) 132:104298.
21. Hsieh C, Zheng K, Lin C, Mei L, Lu L, Li W, et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun*. (2021) 12:5472. doi: 10.1038/s41467-021-25779-x
22. Chang E, Nickel B, Binkley N, Bernatz J, Krueger D, Winzenried A, et al. A novel osteoporosis screening protocol to identify orthopedic surgery patients for preoperative bone health optimization. *Geriatr Orthop Surg Rehabil*. (2022) 6:21514593221116413. doi: 10.1177/21514593221116413
23. Bernatz J, Brooks A, Squire M, et al. Osteoporosis is common and undertreated prior to total joint arthroplasty. *J Arthroplasty*. (2019) 34:1347–53. doi: 10.1016/j.arth.2019.03.044
24. Yagi M, King A, Boachie-Adjei O. Characterization of osteopenia/osteoporosis in adult scoliosis: Does bone density affect surgical outcome? *Spine*. (2011) 36:1652–7. doi: 10.1097/BRS.0b013e31820110b4
25. Wagner S, Formby P, Helgeson M, Kang D. Diagnosing the undiagnosed osteoporosis in patients undergoing lumbar fusion. *Spine*. (2016) 41:E1279–83. doi: 10.1097/BRS.0000000000001612
26. Bjerke B, Zarrabian M, Aleem I, et al. Incidence of osteoporosis-related complications following posterior lumbar fusion. *Global Spine J*. (2018) 8:563–9. doi: 10.1177/2192568217743727
27. Chandran M, Ebeling P, Mitchell P, Nguyen T. Executive committee of the Asia Pacific consortium on osteoporosis. harmonization of osteoporosis guidelines: Paving the way for disrupting the status quo in osteoporosis management in the Asia Pacific. *J Bone Miner Res*. (2022) 37:608–15. doi: 10.1002/jbmr.4544
28. Yamamoto N, Shiroshita A, Kimura R, Kamo T, Ogihara H, Tsuge T. Diagnostic accuracy of chest X-ray and CT using artificial intelligence for osteoporosis: Systematic review and meta-analysis. *J Bone Miner Metab*. (2024) 42:483–91. doi: 10.1007/s00774-024-01532-4
29. Jang M, Kim M, Bae S, Lee S, Koh J, Kim N. Opportunistic osteoporosis screening using chest radiographs with deep learning: Development and external validation with a cohort dataset. *J Bone Miner Res*. (2022) 37:369–77. doi: 10.1002/jbmr.4477
30. Sato Y, Yamamoto N, Inagaki N, Iesaki Y, Asamoto T, Suzuki T, et al. Deep learning for bone mineral density and T-score prediction from chest X-rays: A multicenter study. *Biomedicines*. (2022) 10:2323. doi: 10.3390/biomedicines10092323
31. Asamoto T, Takegami Y, Sato Y, Takahara S, Yamamoto N, Inagaki N, et al. External validation of a deep learning model for predicting bone mineral density on chest radiographs. *Arch Osteoporos*. (2024) 19:15. doi: 10.1007/s11657-024-01372-9
32. Cadarette S, Jaglal S, Murray T, McIsaac W, Joseph L, Brown J, et al. Evaluation of decision rules for referring women for bone densitometry by dual-energy x-ray absorptiometry. *JAMA*. (2001) 286:57–63. doi: 10.1001/jama.286.1.57
33. Broder J. *Diagnostic Imaging for the Emergency Physician, Chapter 5 - Imaging the Chest: The Chest Radiograph*. Amsterdam: Elsevier Health Sciences (2011).
34. Li Y, Yan L, Cai S, et al. The prevalence and under-diagnosis of vertebral fractures on chest radiograph. *BMC Musculoskelet Disord*. (2018) 19:235. doi: 10.1186/s12891-018-2171-y
35. Kasai S, Li F, Shiraishi J, Li Q, Doi K. Computerized detection of vertebral compression fractures on lateral chest radiographs: Preliminary results with a tool for early detection of osteoporosis. *Med Phys*. (2006) 33:4664–74. doi: 10.1118/1.2364053
36. He Y, Lin J, Zhu S, Zhu J, Xu Z. Deep learning in the radiologic diagnosis of osteoporosis: A literature review. *J Int Med Res*. (2024) 52:3000605241244754. doi: 10.1177/03000605241244754
37. Dimai H. Use of dual-energy x-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-score, and reference databases. *Bone*. (2017) 104:39–43. doi: 10.1016/j.bone.2016.12.016
38. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *AAAI*. (2017) 31:12. doi: 10.1609/aaai.v31i1.11231
39. MathWorks. *MATLAB Inceptionresnetv2*. (2024). Available online at: <https://www.mathworks.com/help/deeplearning/ref/inceptionresnetv2.html> (accessed July 5, 2024).
40. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
41. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. (1986) 1:307–10. doi: 10.1016/j.ijnurstu.2009.10.001
42. Yamamoto N, Sukegawa S, Yamashita K, Manabe M, Nakano K, Takabatake K, et al. Effects of patient clinical variables in osteoporosis classification using hip X-rays in deep learning analysis. *Medicina*. (2021) 57:846. doi: 10.3390/medicina57080846
43. Schroeder J, Bigolin L, Li T, Chan J, Vachet C, Paine I, et al. Prediction of obstructive lung disease from chest radiographs via deep learning trained on pulmonary function data. *Int J Chron Obstruct Pulmon Dis*. (2021) 15:3455–66. doi: 10.2147/COPD.S279850
44. Yoshida A, Kai C, Futamura H, Oochi K, Kondo S, Sato I, et al. Spirometry test values can be estimated from a single chest radiograph. *Front Med*. (2024) 11:1335958. doi: 10.3389/fmed.2024.1335958
45. Ueda D, Matsumoto T, Yamamoto A, Walston S, Mitsuyama Y, Takita H, et al. A deep learning-based model to estimate pulmonary function from chest x-rays: Multi-institutional model development and validation study in Japan. *Lancet Digit Health*. (2024) 8:e580–8. doi: 10.1016/S2589-7500(24)00113-4
46. Kusunose K, Hirata Y, Yamaguchi N, Kosaka Y, Tsuji T, Kotoku J, et al. Deep learning approach for analyzing chest x-rays to predict cardiac events in heart failure. *Front Cardiovasc Med*. (2023) 10:1081628. doi: 10.3389/fcvm.2023.1081628