



Article

Deep Learning for Bone Mineral Density and T-Score Prediction from Chest X-rays: A Multicenter Study

Yoichi Sato ^{1,2,3,*} , Norio Yamamoto ^{4,5,6} , Naoya Inagaki ⁷ , Yusuke Iesaki ⁸, Takamune Asamoto ³, Tomohiro Suzuki ⁹ and Shunsuke Takahara ¹⁰

- ¹ Department of Orthopedics Surgery, Japan Community Healthcare Organization (JCHO) Tokyo Shinjuku Medical Center, Tokyo 162-8543, Japan
 - ² Department of Orthopedics Surgery, Nagoya University Graduate School of Medicine, Nagoya 464-8550, Japan
 - ³ Department of Orthopedics Surgery, Gamagori City Hospital, Gamagori 443-8501, Japan
 - ⁴ Department of Orthopedics Surgery, Miyamoto Orthopaedic Hospital, Okayama 703-8236, Japan
 - ⁵ Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama 700-8558, Japan
 - ⁶ Systematic Review Workshop Peer Support Group (SRWS-PSG), Osaka 541-0043, Japan
 - ⁷ Department of Orthopedics Surgery, The Jikei University Kashiwa Hospital, Chiba 277-8567, Japan
 - ⁸ Department of Orthopedics Surgery, The National Hospital Organization Nagoya Medical Center, Nagoya 460-0001, Japan
 - ⁹ iSurgery Co., Ltd., Tokyo 103-0012, Japan
 - ¹⁰ Department of Orthopaedics Surgery, Hyogo Prefectural Kakogawa Medical Center, Kakogawa 675-0003, Japan
- * Correspondence: yoichisugar.trauma@gmail.com



Citation: Sato, Y.; Yamamoto, N.; Inagaki, N.; Iesaki, Y.; Asamoto, T.; Suzuki, T.; Takahara, S. Deep Learning for Bone Mineral Density and T-Score Prediction from Chest X-rays: A Multicenter Study. *Biomedicines* **2022**, *10*, 2323. <https://doi.org/10.3390/biomedicines10092323>

Academic Editor: Yu-Te Wu

Received: 30 August 2022

Accepted: 15 September 2022

Published: 19 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Although the number of patients with osteoporosis is increasing worldwide, diagnosis and treatment are presently inadequate. In this study, we developed a deep learning model to predict bone mineral density (BMD) and T-score from chest X-rays, which are one of the most common, easily accessible, and low-cost medical imaging examination methods. The dataset used in this study contained patients who underwent dual-energy X-ray absorptiometry (DXA) and chest radiography at six hospitals between 2010 and 2021. We trained the deep learning model through ensemble learning of chest X-rays, age, and sex to predict BMD using regression and T-score for multiclass classification. We assessed the following two metrics to evaluate the performance of the deep learning model: (1) correlation between the predicted and true BMDs and (2) consistency in the T-score between the predicted class and true class. The correlation coefficients for BMD prediction were hip = 0.75 and lumbar spine = 0.63. The areas under the curves for the T-score predictions of normal, osteopenia, and osteoporosis diagnoses were 0.89, 0.70, and 0.84, respectively. These results suggest that the proposed deep learning model may be suitable for screening patients with osteoporosis by predicting BMD and T-score from chest X-rays.

Keywords: osteoporosis; screening; DXA; BMD; chest X-ray; deep learning; artificial intelligence

1. Introduction

With the population aging and increasing life expectancy, osteoporosis has become a global health issue affecting more than 200 million people worldwide [1]. It is the greatest risk factor for fragility fractures such as vertebral and hip fractures, and affects life prognosis [2–4]. Early diagnosis of osteoporosis through screening is important for the initiation of therapeutic agents and prevention of fragility fractures [5]. The standard examination for osteoporosis screening is the measurement of bone mineral density (BMD) using dual-energy X-ray absorptiometry (DXA) [6]. However, DXA has drawbacks in terms of high equipment cost and radiation exposure [7–9]. Meanwhile, increasing awareness of osteoporosis may be the most effective strategy for the prevention of osteoporotic

fractures [10]. However, awareness of this disease among the elderly is very low [11]; therefore, the osteoporosis screening rate in Japan is only 5% [12]. Solutions to these challenges would include (1) using commonly available imaging equipment and (2) using multipurpose imaging equipment that is frequently used in clinical settings. Presently, the most frequently performed imaging technique is chest radiography. A previous study demonstrated that values obtained by analyzing structural and anatomical phenotypes, such as the cortical thickness of the clavicles, ribs, and spine in chest radiographs, are correlated with BMD [13–15]. These findings suggest that a tool for predicting BMD from chest X-rays obtained for various medical purposes would be useful for osteoporosis screening. The utilization of chest X-rays taken for other medical purposes eliminates exposure to additional radiation and allows screening without requiring additional medical procedures, specifically for examining bone density.

In recent years, deep learning, which is a machine learning technique that uses multi-layer neural networks, has emerged as an effective technique for improving the performance of computer image recognition [16]. Subsequently, progress in orthopedics research has led to the use of deep learning models for osteoporosis screening [17]. In previous studies, diagnoses of osteoporosis based on radiographs of the lumbar spine and hip joint have been demonstrated [18,19], whereas the BMDs (g/cm^2) of the hip and lumbar spine have been measured from radiographs of these sites [20,21]. Chest X-rays have also been used in two studies to diagnose osteoporosis [22,23]. However, the studies that used chest X-rays predicted only a “T-score less than -2.5 ” or “young adult mean (YAM) less than 80%.” The predictive performance indices obtained from these studies, in terms of the area under the curve (AUC), were 0.88 and 0.78, respectively, and prediction of the BMD (g/cm^2), which is a continuous variable, was not performed. In addition, diagnosis (normal, osteopenia, or osteoporosis) could not be predicted based on the T-score using a single deep learning model. The deep learning models in past studies were trained on datasets that were each obtained from a single site, making it difficult to ensure the validity of the results because of the possibility of overtraining [24]. While a previous study on hip radiographs reported that ensemble learning of image data and patient clinical covariates can increase the prediction accuracy [25], there has been no report of such a learning method based on chest X-rays.

We hypothesized that a deep learning model can predict BMD using chest radiography. The purpose of this study was to develop a deep learning model trained on a large dataset collected from multiple institutions to predict BMD (g/cm^2) and diagnosis based on the T-score (normal, osteopenia, and osteoporosis) using chest X-rays, age, and sex. By developing models with good predictive performance, we may be able to utilize chest X-rays as a screening tool for osteoporosis.

2. Materials and Methods

2.1. Patient Registration and Patient Data Collection

We conducted this retrospective multicenter study by collecting medical data from six hospitals in Japan (one university hospital and five general hospitals). This retrospective study was approved by the ethics committee of the lead hospital. This machine learning-based study was conducted in accordance with the Transparent Reporting of a Multivariable Prediction Model For Individual Prognosis Or Diagnosis (TRIPOD) guidelines [26] (Supplementary File S1).

The inclusion criteria were patients aged 20 years or older who visited any one of the facilities between April 2010 and July 2021, and underwent bone densitometry and chest X-ray imaging. The time gap between the bone densitometry examination and chest X-ray acquisition was within 6 months, in accordance with a previous study [20]. The dataset also included patients with implants or clinical features due to disease within the imaging range of the chest X-rays. The exclusion criteria were as follows: (i) patients whose chest X-rays did not include both lungs and clavicles, and (ii) patients whose chest X-rays were taken using portable equipment.

We extracted the anonymized image files from the image servers. All image files were in the “.dcm” format. The areal BMD was measured at the lumbar spine, femoral neck, and total hip using DXA. The details of the X-ray generator, image processing unit, image size, and DXA scanner at each facility are listed in Supplementary Table S1. We used the patient clinical covariates (age and sex), imaging data (chest X-rays), and results of bone densitometry (BMD and T-score from DXA) for the analysis in this study.

2.2. Data Preparation

We paired the bone densitometry and chest X-ray results of each patient. To improve predictability, we associated the age and sex with the chest X-rays and trained a deep learning model using ensemble learning [25]. We used the BMDs (g/cm^2) measured at (i) the lumbar spine (average of L1–L4) and (ii) the lower of the values measured at the femoral neck and entire proximal femur [9]. The BMD values from the GE scanner were converted to Hologic values using the equations provided in Supplementary Table S2. For the T-score, we used the lowest value of the test results for the lumbar spine (average of L1–L4), femoral neck, and entire proximal femur [9]. We classified the participants into normal, osteopenia, and osteoporosis groups according to the World Health Organization (Geneva) (WHO) criteria [27]. The WHO defines normal as a T-score above -1.0 , osteopenia as a T-score between -1.0 and -2.5 , and osteoporosis as a T-score below -2.5 . We labeled the BMD (g/cm^2) and diagnosis based on the T-score (normal, osteopenia, osteoporosis) calculated against the chest X-rays.

2.3. Splitting the Dataset

We randomly split the dataset collected from each hospital into training, validation, and testing datasets. We ensured that the data for each of the three labels (normal, osteopenia, and osteoporosis), in conjunction with their corresponding chest radiographs, age, and sex ratios, were randomly distributed in balanced numbers among the training, validation, and test datasets. The splitting ratios for the training, validation, and test datasets were 70%, 10%, and 20%, respectively. Figure 1 shows a flowchart of the dataset creation process, which ensured that the test dataset contained only new chest X-ray images that the model did not encounter during the training.

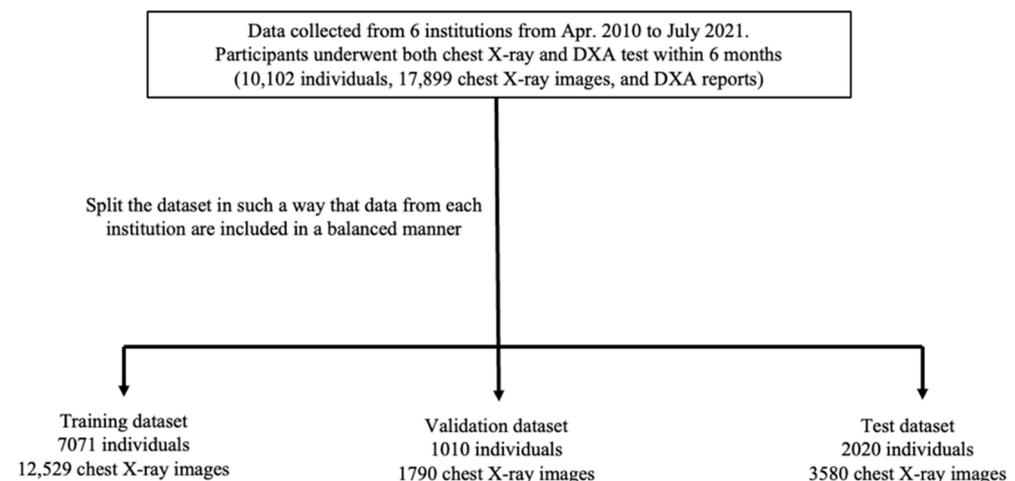


Figure 1. Dataset configuration. Note: DXA: dual-energy X-ray absorptiometry.

2.4. Image Preprocessing and Machine Learning

The specifications of the development environment were as follows: CPU: AMD EPYC 7452, GPU NVIDIA GTX TITAN X, Python 3.8.10, and PyTorch 1.10.0. To improve predictability, we performed data augmentation on the images extracted from the image server. For data augmentation, the image data were amplified via the application of

ColorJitter (random brightness, contrast, saturation, and hue changes), RandomAffine (random geometric deformation), and RandomHorizontalFlip (random left-right flip) to each image. We then decomposed all chest X-rays into four (2×2) patches and resized them to 224×224 pixels. Each decomposed patch was vectorized and concatenated using ResNet50 [28]. These were then combined with the age and sex, and input into a three-layer perceptron with 128 hidden channels. The input batch size was 64 and optimization was performed using stochastic gradient descent. We trained the deep learning model as a regression for BMD and multiclass classification (one-vs.-all classification) for the T-score. In the multi-classification, we trained the deep learning model for three classification tasks as follows: (1) T-score above -1.0 vs. the rest; (2) T-score between -1.0 and -2.5 vs. the rest; and (3) T-score below -2.5 vs. the rest.

2.5. Statistical Analysis

The predictive performance of the deep learning model was evaluated using Scikit-Learn (<https://scikit-learn.org/stable/>; accessed on 1 July 2021). Data analysis was performed using a complete case dataset.

2.5.1. Regression of BMD

We used the Pearson correlation coefficient (R-value), coefficient of determination (R-squared or R^2), and mean absolute error (MAE) as the measures of performance in predicting BMD. In addition, a linear fitting curve and Bland–Altman plots were drawn. The R-value measures the linear correlation between the predicted value and ground truth, and considers only the sequential correlation, regardless of the absolute values. The linear fitting curve illustrates the overall direction of correspondence and modeling quality. The MAE is the error between the predicted values and standard references. Estimating the quality of the methods used for regression requires validation of the correlation between the measured values and reliable standards for accuracy, which is determined through the MAE and standard deviation of the MAE. The linear fitting curve illustrates the overall direction of correspondence and modeling quality. In the Bland–Altman plots, the error is plotted against the average value of a pair of predicted and true values.

2.5.2. Classification of T-Score

The following metrics were used as a measure of performance in the classification of the T-score: (1) accuracy, (2) sensitivity, (3) specificity, and (4) AUC. The 95% confidence interval (CI) was also evaluated. The confusion matrix in this study was set as a 2×2 contingency table displaying the number of true positives, false positives, false negatives, and true negatives. The receiver operating characteristic (ROC) curve was created based on a plot of the true positive rate (sensitivity) against the false positive rate ($1 - \text{sensitivity}$).

3. Results

3.1. Patient Characteristics

The images were chest radiographs of 17,899 individuals (15,060 females and 2839 males, with ages ranging from 24 to 98 years (mean age 71.57 years)). From the chest radiographs, 3152 were categorized as normal results, 10,404 as osteopenia, and 4343 as osteoporosis based on DXA examination. Table 1 presents the baseline characteristics of the training, validation, and testing datasets.

Table 1. Demographic characteristics of the dataset.

		Training Dataset	Validation Dataset	Test Dataset	Overall
Participant		12,529	1790	3580	17,899
age (years), mean \pm SD		71.94 \pm 10.05	71.24 \pm 10.94	71.54 \pm 11.27	71.57 \pm 10.75
Sex	Female (%)	10,544 (84.16%)	1508 (84.25%)	3008 (84.02%)	15,060 (84.14%)
	Male (%)	1985 (15.84%)	282 (15.75%)	572 (15.98%)	2839 (15.86%)
BMD (g/cm ²), mean \pm SD	Lumbar	0.88 \pm 0.19	0.89 \pm 0.21	0.88 \pm 0.20	0.88 \pm 0.20
	Hip	0.58 \pm 0.12	0.59 \pm 0.15	0.58 \pm 0.13	0.58 \pm 0.13
T-score mean \pm SD	Lumbar	-1.51 \pm 1.56	-1.53 \pm 1.68	-1.51 \pm 1.60	-1.52 \pm 1.61
	Hip	-2.145 \pm 1.17	-2.15 \pm 1.40	-2.16 \pm 1.10	-2.15 \pm 1.22
T-score categories, <i>n</i> (%)	Normal	2204 (17.59%)	317 (17.71%)	631 (17.63%)	3152 (17.61%)
	Osteopenia	7287 (58.16%)	1038 (57.99%)	2079 (58.07%)	10,404 (58.13%)
	Osteoporosis	3038 (24.25%)	435 (24.30%)	870 (24.30%)	4343 (24.26%)

Note: BMD: bone mineral density; SD: standard deviation.

3.2. Predictive Performance of Deep Learning Model

3.2.1. Regression of BMD

The correlation plot and Bland–Altman plots for BMD predicted by the deep learning model and true BMD are shown in Figure 2. The predictive performance indices for femoral BMD are as follows: R-value of 0.75, R² of 0.54, and MAE of 0.08. The predictive performance indices for lumbar spine BMD are as follows: R-value of 0.63, R² of 0.40, and MAE of 0.12.

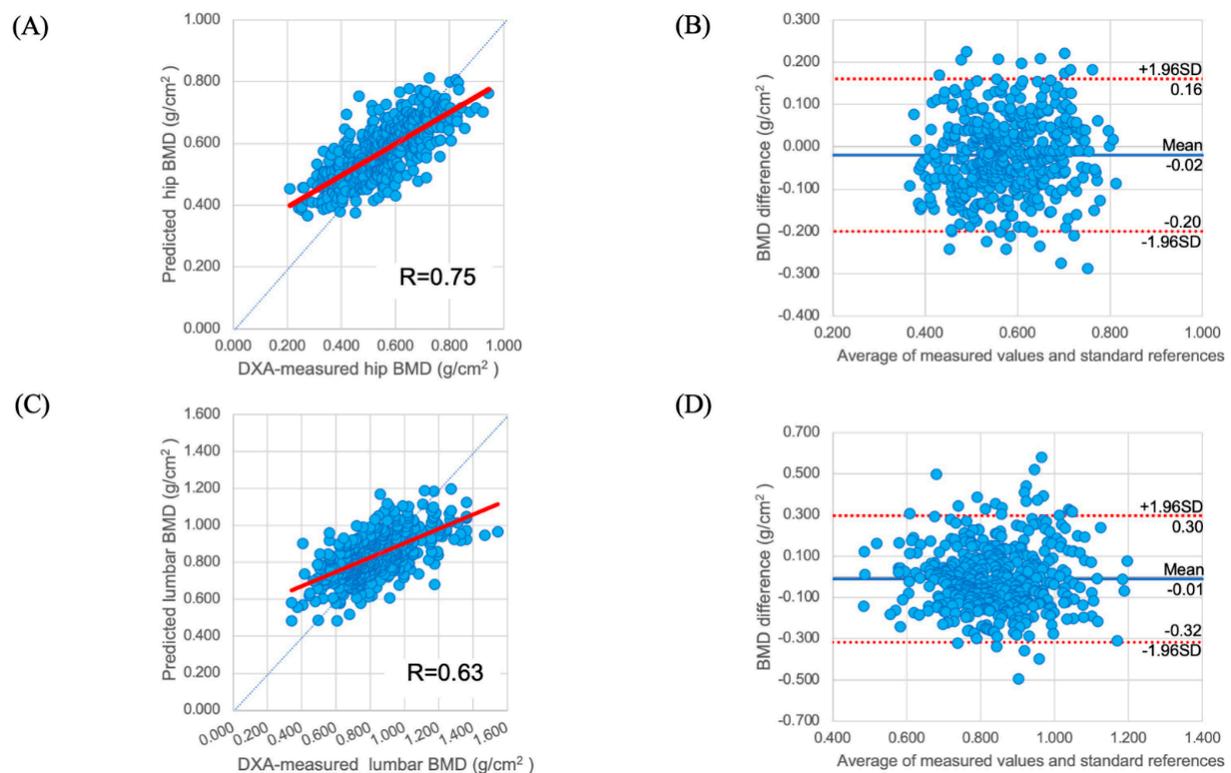


Figure 2. Predictive performance of regression on BMD. (A) Linear fitting curve and (B) Bland–Altman plot for the trained model for femoral BMD prediction. (C) Linear fitting curve and (D) Bland–Altman plot for the trained model for lumbar spine BMD prediction. Model predictions were compared with the ground truth. In the linear fitting curve, R is the Pearson correlation coefficient. Each point in the Bland–Altman plot represents a pair of DXA BMD and predicted BMD; the horizontal axis depicts the mean, whereas the vertical axis depicts the difference. Note: BMD: bone mineral density; DXA: dual-energy X-ray absorptiometry; SD: standard deviation.

3.2.2. Classification of T-Score

The predictive performance of multiclass classification of the diagnoses based on the T-score (normal, osteopenia, and osteoporosis) is shown in Table 2. The ROC curves for the multiclass classification of the T-scores are shown in Figure 3. The predictive performance indices for diagnosis as normal (T-score above -1.0 vs. the rest) are an AUC of 0.89 (95% CI: 0.86–0.91), accuracy of 74.89% (95% CI: 71.21–77.45), sensitivity of 90.14% (95% CI: 87.35–92.41), and specificity of 72.24% (95% CI: 68.32–75.80). The predictive performance indices for diagnosis as osteopenia (T-score between -1.0 and -2.5 vs. the rest) are an AUC of 0.70 (95% CI: 0.68–0.72), accuracy of 66.06% (95% CI: 63.65–68.39), sensitivity of 71.28% (95% CI: 69.01–73.53), and specificity of 62.35% (95% CI: 59.94–64.77). The predictive performance indices for diagnosis as osteoporosis (T-score below -2.5 vs. the rest) are an AUC of 0.84 (95% CI: 0.82–0.86), accuracy of 77.83% (95% CI: 75.52–79.9), sensitivity of 77.27% (95% CI: 74.94–79.36), and specificity of 78.58% (95% CI: 76.32–80.55).

Table 2. Performance metrics of the model for the test dataset. The accuracy, sensitivity, specificity, and AUC in the respective ranges, with T-scores of -1.0 and -2.5 as cutoffs, are shown.

	AUC (95% CI)	Accuracy (%) (95% CI)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)
T-score ≥ -1.0	0.89 (0.86–0.91)	74.89 (71.21–77.45)	90.14 (87.35–92.41)	72.24 (68.32–75.80)
$-1.0 > \text{T-score} > -2.5$	0.70 (0.68–0.72)	66.06 (63.65–68.39)	71.28 (69.01–73.53)	62.35 (59.94–64.77)
$-2.5 \geq \text{T-score}$	0.84 (0.82–0.86)	76.47 (75.52–79.90)	81.25 (74.94–79.36)	73.68 (76.32–80.65)

Note: AUC: area under the curve; CI: confidence interval.

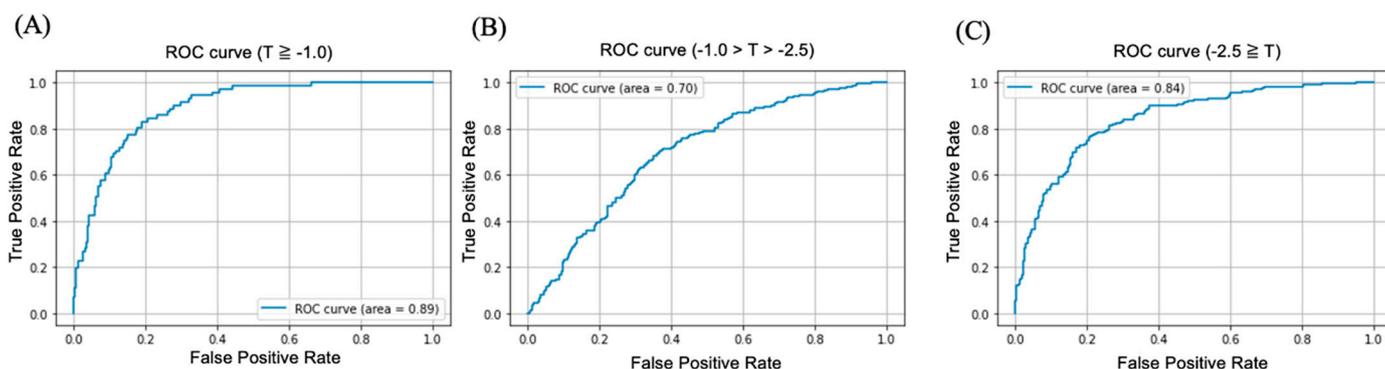


Figure 3. Receiver operating characteristic curve (ROC) for each class: (A) T-score greater than or equal to -1.0 ; (B) T-score between -1.0 and -2.5 ; (C) T-score less than or equal to -2.5 .

4. Discussion

In this study, we developed a deep learning model with ensemble learning based on chest X-rays, age, and sex to predict BMD (g/cm^2) and diagnosis as per the T-score (normal, osteopenia, osteoporosis). With regard to the performance, the deep learning model could predict femoral BMD with $R = 0.75$, and predict “T-score = -1.0 or not” with an AUC of 0.89 and sensitivity of 90.14%. This study is the first to develop a deep learning model that predicts BMD (g/cm^2) and T-scores using multiclass classification based on chest X-rays. The results demonstrated that the deep learning model may have potential for application in osteoporosis screening using chest X-rays in actual clinical practice.

The deep learning model was able to predict BMD using the chest X-rays. The predictive performance for hip BMD was $R = 0.75$, which indicates a high positive correlation with the true value [29]. Because none of the previous studies that predicted osteoporosis from chest X-rays were able to predict BMD, this study represents significant progress in this research area [22,23]. In comparison with the results of studies that predicted BMD from radiographs of the hip and lumbar spine using deep learning models [18,19], the results of our study were slightly inferior (previous studies: $R = 0.81, 0.89$; this study: $R = 0.75$). This may be due to the following reasons. (1) The site corresponding to the radiograph and

the site where BMD was measured were different. (2) The training was performed based on setting the region of interest of the bone or dividing the image into sections instead of considering the entire image. Based on these factors, training the learning model such that the lumbar spine is cut out from the chest X-rays may improve the predictive performance. However, the performance of the method cannot be guaranteed. A previous study reported less accurate results in predicting the BMD of the lumbar spine than in predicting that of the hip [19]. Similarly, in this study, the predictive performance of the BMD differed between the hip and lumbar spine (hip: $R = 0.75$; lumbar spine: $R = 0.63$). The reason for this may be that, in comparison with those at the hip joint, the DXA measurements at the lumbar spine are subject to measurement errors due to osteoarthritis [30]. To address this problem, it is necessary to verify whether the performance can be improved through modifications of the labels and reorganization of the dataset.

The deep learning model was also able to predict diagnosis with moderate performance by utilizing T-scores with multiclass classification (normal, osteopenia, and osteoporosis) based on chest X-rays. The predictive performance indices were AUC = 0.89, 0.70, and 0.84, respectively. The predictive performance in the diagnoses of normal and osteopenia could not be compared because of the absence of similar studies in literature, but the predictive performance in diagnosis of osteoporosis was slightly inferior to that of a previous study [23]. Compared with previous studies that diagnosed osteoporosis using chest X-rays [22,23], our study has the following novel aspects: (1) a single deep learning model is classified into three classes: normal, osteopenia, and osteoporosis (multiclass classification); and (2) the T-score is used to predict diagnosis (normal or osteopenia). In screening for osteoporosis, it is important not only to identify the participants with T-scores below -2.5 , but also those with T-scores between -1.0 and -2.5 . This is because among the participants who underwent bone densitometry, the group diagnosed with osteoporosis had a higher fracture rate, whereas the group diagnosed with osteopenia had a higher number of patients. Therefore, the total number of fractures was higher in the group diagnosed with osteopenia than in the group diagnosed with osteoporosis [31]. Medical guidelines recommend further examination or therapeutic interventions for osteopenia [9,32,33]. Therefore, a deep learning model that can identify osteopenia is necessary. With regard to the predictive performance for T-score = -2.5 , it was slightly lower in this study than in previous studies (Jang et al. [23]: AUC = 0.88; this study: AUC = 0.84). This was because in this study, data were collected from multiple centers, and thus a broad range of inclusion criteria was set. Large-scale and comprehensive data collection is necessary to ensure versatility. The previous study cited these factors as limitations, which were overcome in this study. The inferior performance indicates that there is potential for performance improvement. Previous studies have reported that learning based on setting regions of interest (shoulder, cervical and thoracic area, thoracic, and lumbar area) in chest X-rays improves the performance [23]. In the future, we will train our model using this approach.

Our deep learning model has the potential to perform osteoporosis screening using chest X-rays. In Japanese osteoporosis screening, a T-score below -1.0 indicates that the patient needs further examination. The predictive performance indices of the deep learning model developed in this study, with T-score = -1.0 as the cutoff, were sensitivity = 90.14% and specificity = 72.24%. From the viewpoint of triage screening for osteoporosis, high sensitivity (approximately 90%) and relatively low specificity (approximately 40–60%) are considered acceptable for clinical decision rules [34]. Therefore, we can use this deep learning model to screen for osteoporosis. In Japan, 40 million people over the age of 40 are screened for lung cancer using chest X-rays [35]. By applying the deep learning model to these potential participants to screen for osteoporosis, we could find five million new osteoporosis patients based on the age range of the examinees and age-specific incidence of osteoporosis [35,36]. Appropriate therapeutic interventions for these patients would then help prevent fragility fractures [37].

The strength of this study lies in the collection of diverse data from multiple institutions. The advantages of multicenter studies are (i) the ability to prevent overfitting by collecting a large amount of data [24] and (ii) the ability to conduct comprehensive research by using data obtained from different conditions and environments, thereby allowing medical research to be conducted in clinical settings [38]. In this study, we collected approximately 18,000 training data points from approximately 10,000 cases, which included almost all chest X-ray images taken at multiple institutions in Japan and with various medical devices over a long duration. This allowed for diverse patient datasets (images that included implants and clinical features due to disease) collected from multiple examination settings, including X-ray generators, image processing units, and DXA scanners. This supports this study's validity as an epidemiological study and ensures its internal validity. On the contrary, to be used in clinical practice, external validity must be assessed using data from other institutions.

However, this study has several limitations. First, we did not develop multiple trained models or validate their predictive performance. Transfer learning using pretrained models is common in deep learning [39]. A previous study evaluated various learned models and reported differences in their performance [20]. In this study, we used ResNet50 because of its short processing time [28]. In the future, training with different learning models may lead to improved performance. Second, we considered only age and sex as the patient variables in predicting the BMD and T-scores. However, various patient factors can influence the incidence of osteoporotic fractures [40]. In this study, we trained our deep learning model using chest X-rays, age, and sex. This was because we believed that learning from the information contained in the image file (image, age, and sex) would not change the current workflow in an actual clinical setting. However, a previous study reported that training a deep learning model with patient clinical covariates, such as height, weight, and fracture history, improved the performance [25]. Further, various diseases (COPD, rheumatism, etc.) that coexist with osteoporosis should be considered [41,42]. Considering this, we can train our deep learning model with these factors to verify the possibility of improving the performance. Third, we have not evaluated the predictive accuracy of the developed training model for each age group (young, middle-aged, and older adults). Osteoporosis is prevalent in aged women, and this population group is the target for screening [9]. Secondary analysis for this age group is required to make the analysis more relevant to actual clinical practice. Fourth, we did not perform an external validation. Most studies on deep learning models have not evaluated the validity of the models in different environments [38]. Although this study prepared a dataset with data collected from multiple facilities, we were unable to validate our model using data from entirely different clinical settings. To train our deep learning model as a programmed medical tool, it is necessary to evaluate the predictive performance using data collected at different facilities and from different racial groups. Fifth, while the deep learning model could diagnose osteoporosis on guidelines based on T-score, this did not necessarily imply that it could understand the pathophysiology of osteoporosis, including causative disease and comorbidities. We developed this deep learning model using radiographs, bone densitometry, age, and gender but did not consider medical history such as comorbidities. Therefore, to confirm whether the results of this deep learning model analysis are normal, the physician should interview and examine the patient, perform blood tests, and make a definitive diagnosis using DXA.

5. Conclusions

We developed a deep learning model based on ensemble learning of chest X-rays, age, and sex to predict BMD (g/cm^2) and diagnosis according to the T-score (normal, osteopenia, osteoporosis). With this model, chest X-rays taken for various medical reasons can be used to identify patients at risk for osteoporosis without additional radiation exposure or cost, and without the possibility of behavioral changes in the examinee. This may improve

screening for osteoporosis. To realize the goal of clinical application, we need to further improve the predictive performance and validity of the deep learning model.

6. Patents

A patent application for the results of this study has been filed in Japan (No. 21ZP324).

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines10092323/s1>, Table S1: title X-ray generator, image processing unit, image size, and DXA scanner at each facility; Table S2: title Equations for converting BMD value of GE to that of Hologic DXA scanners.

Author Contributions: Conceptualization, Y.S.; Methodology, Y.S., N.Y. and S.T.; Software, T.S.; Validation, Y.S. and T.S.; Formal analysis, T.S.; Investigation, Y.S.; Resources, Y.S., N.Y., S.T., N.I. and Y.I.; Data curation, Y.S.; Writing—original draft preparation, Y.S.; Writing—review and editing, N.Y. and S.T.; Visualization, Y.S.; Supervision, S.T.; Project administration, Y.S. and T.A.; Funding acquisition, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JOA = Subsidized Science Project Research 2022-1, Research Grants of Mitsui Sumitomo Insurance Welfare Foundation 2021, and iSurgery Co., Ltd.

Institutional Review Board Statement: This retrospective study was approved by the ethics committee of the lead hospital (JCHO Tokyo Shinjuku Medical Center: IRB No. R3-14). This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki and the current scientific guidelines.

Informed Consent Statement: This was a retrospective observational study dealing with anonymized processed information. Consent was obtained on an opt-out basis following Japanese ethical regulations, the Japanese Personal Information Protection Law, and the instructions of each ethics committee. In the opt-out method, information regarding the purpose and methodology of the study was notified or disclosed, and opportunities for refusal were guaranteed to the extent possible.

Data Availability Statement: The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available because of privacy or ethical restrictions.

Acknowledgments: We are grateful to Seiwa Honda, Atsuo Uefuji, Ryota Nishida, Koki Toge, Takehiro Konishi, Ryo Okada, Takahiro Kitano, Saya Goto, Katsuyuki Isozaki, and Shiina Ichimori for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sözen, T.; Özişik, L.; Başaran, N.Ç. Overview and management of osteoporosis. *Eur. J. Rheumatol.* **2017**, *4*, 46–56. [[CrossRef](#)] [[PubMed](#)]
2. Suzuki, T.; Yoshida, H. Low bone mineral density at the femoral neck is a predictor of increased mortality in elderly Japanese women. *Osteoporos. Int.* **2010**, *21*, 71–79. [[CrossRef](#)] [[PubMed](#)]
3. Ensrud, K.E.; Thompson, D.E.; Cauley, J.A.; Nevitt, M.C.; Kado, D.M.; Hochberg, M.C.; Santora, A.C., II; Black, D.M. 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*, 241–249. [[CrossRef](#)] [[PubMed](#)]
4. Nguyen, N.D.; Center, J.R.; Eisman, J.A.; Nguyen, T.V. Bone loss, weight loss, and weight fluctuation predict the mortality risk in elderly men and women. *J. Bone Miner Res.* **2007**, *22*, 1147–1154. [[CrossRef](#)] [[PubMed](#)]
5. Mitchell, P.J. Fracture Liaison Services: UK Experience. *Osteoporos. Int.* **2011**, *22* (Suppl. S3), 487–494. [[CrossRef](#)]
6. US Preventive Services Task Force. Clinical guideline: Screening for osteoporosis: U.S. preventive services task force recommendation statement. *Encycl. Ann. Intern. Med.* **2011**, *154*, 356–364. [[CrossRef](#)]
7. Mueller, D.; Gandjour, A. Cost-effectiveness of using clinical risk factors with and without DXA for osteoporosis screening in postmenopausal women. *Value Health* **2009**, *12*, 1106–1117. [[CrossRef](#)]
8. Sim, M.F.V.; Stone, M.; Johansen, A.; Evans, W. Cost effectiveness analysis of BMD referral for DXA using ultrasound as a selective pre-screen in a group of women with low trauma colle fractures. *Technol. Health Care* **2000**, *8*, 277–284. [[CrossRef](#)]
9. Orimo, H.; Nakamura, T.; Hosoi, T.; Iki, M.; Uenishi, K.; Endo, N.; Ohta, H.; Shiraki, M.; Sugimoto, T.; Suzuki, T.; et al. Japanese 2011 guidelines for prevention and treatment of osteoporosis—executive summary. *Arch. Osteoporos.* **2012**, *7*, 3–20. [[CrossRef](#)]

10. Sedlak, C.A.; Doheny, M.O.; Jones, S.L. Osteoporosis education programs: Changing knowledge and behaviors. *Public Health Nurs.* **2000**, *17*, 398–402. [[CrossRef](#)]
11. Sato, M.; Vietri, J.; Flynn, J.A.; Fujiwara, S. Bone fractures and feeling at risk for osteoporosis among women in Japan: Patient characteristics and outcomes in the National Health and Wellness Survey. *Arch. Osteoporos.* **2014**, *9*, 199. [[CrossRef](#)] [[PubMed](#)]
12. Taguchi, A. Triage screening for osteoporosis in dental clinics using panoramic radiographs. *Oral Dis.* **2010**, *16*, 316–327. [[CrossRef](#)]
13. Kumar, D.A.; 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. Investig.* **2014**, *37*, 429–440. [[CrossRef](#)] [[PubMed](#)]
14. Chen, H.; Zhou, X.; Fujita, H.; Kubo, K.Y. Age-related changes in trabecular and cortical bone microstructure. *Int. J. Endocrinol.* **2013**, *2013*, 213234. [[CrossRef](#)] [[PubMed](#)]
15. Holcombe, S.A.; Hwang, E.; Derstine, B.A.; Wang, S.C. Measurement of rib cortical bone thickness and cross-section using CT. *Med. Image Anal. Elsevier B.V.* **2018**, *49*, 27–34. [[CrossRef](#)] [[PubMed](#)]
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on image classification. *Proc. IEEE Int. Conf. Comput. Vis.* **2015**, *2015*, 1026–1034.
17. Smets, J.; Shevroja, E.; Hügle, T.; Leslie, W.D.; Hans, D. Machine learning solutions for osteoporosis—a review. *J. Bone Miner Res.* **2021**, *36*, 833–851. [[CrossRef](#)]
18. Nguyen, T.P.; Chae, D.S.; Park, S.J.; Yoon, J. A 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. [[CrossRef](#)]
19. Hsieh, C.I.; Zheng, K.; Lin, C.; Mei, L.; Lu, L.; Li, W.; Chen, F.-P.; Wang, Y.; Zhou, X.; Wang, F.; et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat. Commun.* **2021**, *12*, 5472. [[CrossRef](#)]
20. Yamamoto, N.; Sukegawa, S.; Kitamura, A.; Goto, R.; Noda, T.; Nakano, K.; Takabatake, K.; Kawai, H.; Nagatsuka, H.; Kawasaki, K.; et al. Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates. *Biomolecules* **2020**, *10*, 1534. [[CrossRef](#)]
21. Zhang, B.; Yu, K.; Ning, Z.; Wang, K.; Dong, Y.; Liu, X.; Wang, J.; Zhu, C.; Yu, Q.; Duan, Y.; et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study. *Bone* **2020**, *140*, 115561. [[CrossRef](#)] [[PubMed](#)]
22. Ohta, Y.; Yamamoto, K.; Matsuzawa, H.; Kobayashi, T. Development of a fast screening method for osteoporosis using chest X-ray images and machine learning. *Can. J. Biomed. Res. Technol.* **2020**, *3*, 3–9.
23. Jang, M.; Kim, M.; Bae, S.J.; Lee, S.H.; Koh, J.M.; 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*, 369–377. [[CrossRef](#)] [[PubMed](#)]
24. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [[CrossRef](#)]
25. Yamamoto, N.; Sukegawa, S.; Yamashita, K.; Manabe, M.; Nakano, K.; Takabatake, K.; Kawai, H.; Ozaki, T.; Kawasaki, K.; Nagatsuka, H.; et al. Effects of patient clinical variables on osteoporosis classification using hip X-rays in deep learning analysis. *Medicina* **2021**, *57*, 846. [[CrossRef](#)] [[PubMed](#)]
26. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): TRIPOD statement. *BMC Med.* **2015**, *13*, 1. [[CrossRef](#)] [[PubMed](#)]
27. Dimai, H.P. Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-scores, and reference databases. *Bone Elsevier Inc.* **2017**, *104*, 39–43. [[CrossRef](#)]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern. Recognit.* **2016**, *2016*, 770–778.
29. Mukaka, M.M. Statistics corner: A guide to the appropriate use of correlation coefficients in medical research. *Malawi Med. J.* **2012**, *24*, 69–71.
30. Liu, G.; Peacock, M.; Eilam, O.; Dorulla, G.; Braunstein, E.; Johnston, C.C. Effect of osteoarthritis in the lumbar spine and hip on bone mineral density and osteoporosis diagnosis in elderly men and women. *Osteoporos. Int.* **1997**, *7*, 564–569. [[CrossRef](#)]
31. Siris, E.S.; Chen, Y.T.; Abbott, T.A.; Barrett-Connor, E.; Miller, P.D.; Wehren, L.E.; Berger, M.L. Bone mineral density thresholds for pharmacological interventions to prevent fractures. *Arch. Intern. Med.* **2004**, *164*, 1108–1112. [[CrossRef](#)] [[PubMed](#)]
32. Kanis, J.A.; Cooper, C.; Rizzoli, R.; Reginster, J.Y. Scientific Advisory Board of the European Society for Clinical and Economic Aspects of Osteoporosis (ESCEO) and the Committees of Scientific Advisors and National Societies of the International Osteoporosis Foundation (IOF). Correction to: European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos. Int.* **2020**, *31*, 209. [[PubMed](#)]
33. Camacho, P.M.; Petak, S.M.; Binkley, N.; Diab, D.L.; Eldeiry, L.S.; Farooki, A.; Harris, S.T.; Hurley, D.L.; Kelly, J.; Lewiecki, M.; et al. American Association of Clinical Endocrinologists/American College of Endocrinology Clinical Practice Guidelines for The Diagnosis and Treatment of Postmenopausal Osteoporosis-2020 Update. *Endocr. Pract.* **2020**, *26* (Suppl. S1), 1–46. [[CrossRef](#)]
34. Cadarette, S.M.; Jaglal, S.B.; Murray, T.M.; McIsaac, W.J.; Joseph, L.; Brown, J.P. Evaluation of decision rules for referring women for bone densitometry by dual-energy X-ray absorptiometry. *J. Am. Med. Assoc.* **2001**, *286*, 57–63. [[CrossRef](#)] [[PubMed](#)]
35. Cancer Information Service. Available online: https://ganjoho.jp/reg_stat/index.html (accessed on 16 May 2022).

36. Yoshimura, N.; Muraki, S.; Oka, H.; Mabuchi, A.; En-Yo, Y.; Yoshida, M.; Saika, A.; Yoshida, H.; Suzuki, T.; Yamamoto, S.; et al. Prevalence of knee osteoarthritis, lumbar spondylosis, and osteoporosis in Japanese men and women: The research on osteoarthritis/osteoporosis against disability study. *J. Bone Miner Metab.* **2009**, *27*, 620–628. [[CrossRef](#)]
37. Cummings, S.R.; Black, D.M.; Thompson, D.E.; Applegate, W.B.; Barrett-Connor, E.; Musliner, T.A.; Palermo, L.; Prineas, R.; Rubin, S.M.; Scott, J.C.; et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: Results from the Fracture Intervention. *Trial. J. Am. Med. Assoc.* **1998**, *280*, 2077–2082. [[CrossRef](#)]
38. Kim, D.W.; Jang, H.Y.; Kim, K.W.; Shin, Y.; Park, S.H. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. *Korean J. Radiol.* **2019**, *20*, 405–410. [[CrossRef](#)]
39. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]
40. Cruz, A.S.; Lins, H.C.; Medeiros, R.V.A.; Filho, J.M.F.; da Silva, S.G. Artificial intelligence on the identification of risk groups for osteoporosis, a general review. *Biomed. Eng. Online* **2018**, *17*, 12. [[CrossRef](#)]
41. Hirano, K.; Imagama, S.; Hasegawa, Y.; Ito, Z.; Muramoto, A.; Ishiguro, N. The influence of locomotive syndrome on health-related quality of life in a community-living population. *Mod. Rheumatol.* **2013**, *23*, 939–944. [[CrossRef](#)]
42. Haugeberg, G.; Uhlig, T.; Falch, J.A.; Halse, J.I.; Kvien, T.K. Bone mineral density and frequency of osteoporosis in female patients with rheumatoid arthritis: Results from 394 patients in the Oslo County rheumatoid arthritis register. *Arthritis Rheum.* **2000**, *43*, 522–530. [[CrossRef](#)]