

Review

# The Diagnostic Value of Image-Based Machine Learning for Osteoporosis: Systematic Review and Meta-Analysis

Rui Zhao<sup>1</sup>, MM; Haolin Yang<sup>1</sup>, MM; Yangbo Li<sup>1</sup>, MM; Xiaoyun Li<sup>1</sup>, MM; Zhijie Yang<sup>1</sup>, MM; Yanping Lin<sup>2</sup>, MM; Jiachun Huang<sup>2</sup>, MD; Lei Wan<sup>2</sup>, MD; Hongxing Huang<sup>2</sup>, MD

<sup>1</sup>The Third Clinical College of Medicine, Guangzhou University of Traditional Chinese Medicine, Guangzhou, China

<sup>2</sup>The Third Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou, China

**Corresponding Author:**

Hongxing Huang, MD  
The Third Affiliated Hospital of Guangzhou University of Chinese Medicine  
Guangzhou University of Chinese Medicine  
261 Longxi Avenue, Liwan District  
Guangzhou 510000  
China  
Phone: 86 13922726488  
Email: [hxh@gzucm.edu.cn](mailto:hxh@gzucm.edu.cn)

## Abstract

**Background:** Osteoporosis (OP) is projected to be a major issue significantly impacting the well-being of middle-aged and old populations. Machine learning (ML) and deep learning (DL) models developed based on medical imaging have enhanced clinicians' diagnostic accuracy and work efficiency. However, the diagnostic performance of different types of medical imaging for OP has not been systematically assessed.

**Objective:** By summarizing related literature, this study aims to elucidate the role of DL models based on different medical imaging modalities in OP detection.

**Methods:** PubMed, Embase, the Cochrane Library, and Web of Science were systematically searched for studies using ML for the diagnosis of OP based on medical imaging. The final search was conducted on May 16, 2024. The risk of bias in the included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. A bivariate mixed-effects model was applied to perform meta-analyses of sensitivity (SEN) and specificity (SPC), stratified by imaging modality (x-ray, computed tomography [CT], magnetic resonance imaging [MRI]). In addition, subgroup analyses were carried out based on the type of ML algorithm, the method of validation dataset generation, and the anatomical site of assessment.

**Results:** A total of 60 studies comprising 66,195 participants were encompassed in this systematic review and meta-analysis. Among these, 22 studies used x-ray imaging, 37 applied CT imaging, and 3 used MRI for ML-based OP diagnosis. For x-ray-based models, the pooled SEN and SPC for studies focusing on the appendicular skeleton were 0.97 (95% CI 0.83-0.99) and 0.90 (95% CI 0.75-0.96), respectively. For studies using the mandible as the target site, SEN and SPC were 0.94 (95% CI 0.89-0.97) and 0.80 (95% CI 0.56-0.93), respectively. For those focusing on the lumbar spine, the pooled SEN and SPC were 0.87 (95% CI 0.77-0.93) and 0.82 (95% CI 0.75-0.87), respectively. For CT-based models, studies targeting the hip joint reported a pooled SEN and SPC of 0.87 (95% CI 0.83-0.90) and 0.92 (95% CI 0.81-0.96), respectively. For the thoracic spine, SEN and SPC were 0.91 (95% CI 0.86-0.94) and 0.94 (95% CI 0.92-0.95), respectively, while for the lumbar spine, they were 0.91 (95% CI 0.87-0.94) and 0.92 (95% CI 0.86-0.95), respectively.

**Conclusions:** ML based on medical imaging demonstrates high diagnosis accuracy for OP, particularly DL models using x-ray and CT modalities. However, this study included only a limited number of original studies using MRI-based ML, and there remains a lack of adequate external validation across studies, which poses interpretative limitations. Future research should aim to develop artificial intelligence tools with broader applicability and enhanced diagnostic precision.

*J Med Internet Res* 2026;28:e75965; doi: [10.2196/75965](https://doi.org/10.2196/75965)

**Keywords:** osteoporosis; machine learning; artificial intelligence; systematic review; diagnostic imaging

## Introduction

Osteoporosis (OP), a metabolic disorder, features a systemic reduction in bone mass and impaired bone micro-architecture and elevates the risk of fragility fractures. As the most prevalent chronic metabolic bone disease, it is strongly associated with advancing age, posing significant health threats. However, due to its insidious onset, prolonged disease course, and challenges in treatment, public awareness and attention toward OP prevention and management remain insufficient [1,2]. With the emerging global trend of population aging, OP is projected to become a major issue adversely affecting the quality of life of middle-aged and older people. Epidemiological studies estimate that by 2050, the global population at high risk of fractures will surge to 6.26 million from 1.66 million in 1990 [3]. This escalation imposes immense social pressures and substantial economic burdens on early OP screening, prevention, and treatment.

At present, a variety of diagnostic methods are available for the clinical assessment of OP. Among them, dual-energy x-ray absorptiometry (DXA) for measuring *T* scores recommended by the World Health Organization is regarded as the authoritative and standardized technique [4]. Although DXA is widely used, it is unable to assess whole-body skeletal, fat, and lean mass, which restricts its utility in the routine diagnosis or evaluation of OP [5]. Moreover, due to disparities in socioeconomic development across different regions worldwide, DXA is not accessible in underdeveloped countries and regions. Therefore, some high-risk populations, such as postmenopausal women and older adults, are not detected and untreated. Medical imaging is crucial in clinical diagnosis and treatment. However, the hidden features within imaging techniques including x-rays, computed tomography (CT), and magnetic resonance imaging (MRI) are often overlooked due to low spatial resolution and high contrast resolution [6].

In the 1980s, computer-aided diagnosis (CAD) systems were developed to deeply interpret key features in medical images, providing radiologists with valuable insights into image interpretation [7]. Currently, CAD tools primarily include traditional machine learning (ML) models built on explainable clinical features and deep learning (DL) models developed using pathological or nuclear medicine images. They assist clinicians in disease diagnosis and prognostic prediction. Increasing evidence has demonstrated the utility of CAD in diagnosing conditions such as autism [8], pulmonary embolism [9], breast cancer [10], and bone metastases [11]. DL approaches based on medical imaging have attracted substantial research interest. Against this backdrop, ML models based on various imaging modalities such as x-rays, CT, and MRI have been constructed to diagnose OP [12]. However, the diagnostic performance of various imaging methods in OP is not supported by systematic evidence. This hindered the application of artificial intelligence (AI)-based CAD tools in OP and posed challenges for further systematic development.

Therefore, our study seeks to provide a comprehensive review of DL research in the diagnosis of OP based on medical imaging modalities, including x-ray, CT, and MRI. Furthermore, this study aims to analyze and evaluate the feasibility and accuracy of AI-driven DL in enhancing the screening and diagnostic rates of OP, thereby offering robust support for the prevention and management of the disease.

## Methods

### Study Registration

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was prospectively registered on PROSPERO (CRD42024567736).

### Eligibility Criteria

The eligible studies were (1) case-control, cohort, or cross-sectional studies; (2) papers with comprehensively developed image-based DL models for OP diagnosis; and (3) English publications. The following studies were excluded: (1) studies that only developed traditional ML models, (2) those that performed image segmentation without a complete DL model, and (3) those lacking outcome measures for evaluating the DL model's accuracy. Outcome measures must include at least 1 of the following: *c*-statistic, sensitivity (SEN), specificity (SPC), accuracy, recall, precision, confusion matrix, *F*<sub>1</sub>-score, or calibration curve.

### Data Sources and Search Strategy

PubMed, Cochrane, Embase, and Web of Science databases were thoroughly retrieved up to May 16, 2024. Both MeSH and free-text terms were used without restrictions on geographic location or study type. The search strategy is detailed in [Multimedia Appendix 1](#).

### Study Selection

The retrieved literature was uploaded to EndNote (Thomson Corporation), and duplicates were ostracized. Titles and abstracts were reviewed to identify potentially eligible studies. Full-text papers were subsequently screened to determine the eligible ones. Two researchers (RZ and HY) independently conducted the literature screening and cross-checked their results. Dissents were addressed by a third researcher (YL).

### Data Extraction

The eligible papers were imported into EndNote, and data extraction was performed. A standard electronic data extraction form was developed beforehand to capture the following information: title, DOI, first author, publication year, author's country, study type, patient source, OP diagnosis criteria, medical imaging, background population, gender, age, use of image segmentation, number of OP cases, total cases, number of OP and total cases in the training or validation set, validation set generation method, model type, and comparison with clinical practitioners. Data were independently extracted by 2 researchers (RZ and HY),

followed by cross-checking. There was a high level of agreement between the 2 reviewers in the screening process (Cohen  $\kappa=0.879$ ). In cases of disagreement, a third reviewer (YL) would assist in addressing it.

### **Risk of Bias in Studies**

The bias of risk in the eligible studies was assessed via Quality Assessment of Diagnostic Accuracy Studies-2, a tool for evaluating the collation risk of bias and clinical applicability of original diagnostic studies [13]. Quality Assessment of Diagnostic Accuracy Studies-2 covers 5 domains: case selection, trials to be evaluated, reference standard, case flow, and progress, with each involving a few specific questions. The answer of “Yes,” “No,” or “Uncertain” corresponds to a low, high, or uncertain risk of bias. The risk of bias was deemed low if all of the landmark questions within a range were answered with “Yes”; if one of the informative questions was answered with “No,” bias may exist, and the evaluators must determine the risk of bias in line with the established guidelines. The risk of bias must be judged by the evaluation authors as per the established criteria. An unclear risk indicated that the studies reported sufficient details. Therefore, evaluators could not make a definitive judgment.

The risk of bias in studies was independently conducted by 2 researchers (RZ and HY), followed by cross-checking. If any dissent arose, a third researcher (YL) would assist in addressing it.

### **Synthesis Methods**

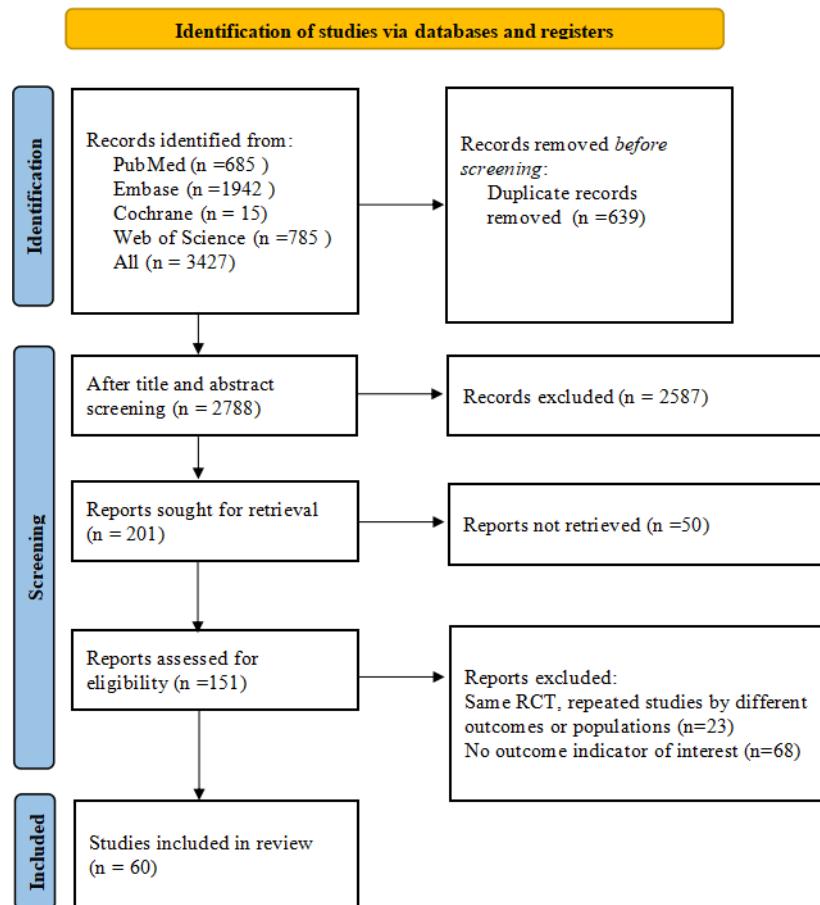
The meta-analysis was carried out via a bivariate mixed-effects model based on diagnostic 2×2 contingency tables. However, some of the original studies did not provide complete 2×2 diagnostic data. In such cases, the necessary information was derived using SEN, SPC, positive predictive

value, negative predictive value, and accuracy, in conjunction with the corresponding sample sizes. The meta-analysis reported pooled estimates of SEN, SPC, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and the summary receiver operating characteristic (SROC) curve along with their corresponding 95% CIs. Publication bias across studies was assessed through Deeks’ funnel plot, while the clinical utility of the predictive models was evaluated via Fagan’s nomogram. Subgroup analyses were performed based on imaging modality (x-ray, CT, and MRI), modeling approach (traditional ML vs DL), and validation strategy. It is important to note that the bivariate mixed-effects model requires a minimum of four 2×2 diagnostic tables. As the ML models based on MRI images only provided 3 such tables, a narrative synthesis was performed for this subgroup instead. A 2-sided  $P$  value of  $<.05$  denoted statistical significance.

## **Results**

### **Study Registration**

A total of 3427 papers were retrieved, including 685 from PubMed, 15 from Cochrane, 1942 from Embase, and 785 papers from Web of Science. Among them, 639 papers were duplicates and were excluded. After the title and abstract review, 2587 studies unrelated to the study topic were removed. Full texts of the rest were subsequently reviewed. In total, 23 conference abstracts without full-text publications and 68 that did not include medical imaging in the modeling process were ostracized. Ultimately, 60 studies were included in the analysis (Figure 1) [14-73]. This study was conducted in accordance with the PRISMA 2020 checklist (Checklist 1).

**Figure 1.** Literature screening process. RCT: randomized controlled trial.

## Study Characteristics

Among the 60 studies included in our analysis, 55 were case-control studies [14-21,25-35,37-67,69-73], and 5 were cohort studies [22-24,36,68]. These studies were predominantly published between 2012 and 2024 and involved 66,195 cases. These studies were published in 11 countries, including China (n=27), South Korea (n=10), the United States (n=8), and Japan (n=5) [15-17,20,22,23,26-33,35-48,50,51,53-60, 62-73]. A smaller number of studies were from India (n=3), Saudi Arabia (n=2), Jordan (n=1), Latvia (n=1), Malaysia (n=1), Poland (n=1), and Switzerland (n=1) [14,18,19,21,24, 25,34,39,52,61]. In total, 57 studies reported patient sources, of which 42 were single-center studies [14,15,17,19-23,26,29, 31,32,36-42,47,48,50-60,62,64-67,69-73], 12 were multicenter studies [16,27,28,30,33,34,43-45,61,63,68], and 4 used database sources [24,41,46,49]. In terms of OP diagnosis, 47 studies explicitly provided diagnosis criteria [16,17,20-23,25, 27-33,35-42,44,45,47,50-58,60-64,66-73]. Regarding medical imaging, 37 studies developed CT-based imaging models [15,16,18,19,21-23,25,26,29,31-33,38-41,43-45,47,48,50- 54,56-58,65,67-69,72,73], 22 developed x-ray-based models [14,17,20,24,27,28,30,34-37,42,46,49,55,59-63,70,71], and 3 focused on MRI-based models [45,64,66]. Concerning the population, 4 studies specifically examined postmenopausal women [16,35,36,52,71], and 1 study focused on men aged 50 years and older [16]. In terms of image processing, 48 studies used manual segmentation techniques to define regions of interest for analysis [14-19,21-23,25-36,38,39,

41,42,44,45,47,49,52-57,59-66,68,70-73], while 12 did not define regions of interest [20,24,37,40,43,46,48,50,51,58,67, 69]. Regarding the skeletal parts, 33 studies used lumbar vertebrae images [15-17,20,22,23,26,29-32,36,38-41,45,47, 48,50,53-55,57,62,64-68,70,71,73], 9 used thoracic vertebrae images [23,25,31,38,44,48,53,55,56], 10 used hip images (including femoral neck, femoral head, and pelvis) [18,21,33, 36,57,62,68,70,72], 7 used mandible images [17,36,42,52,60, 63,71], and 10 used images of limb bones [14,28,34,35,43,51, 58,59,61,69]. Regarding the generation of validation sets, 35 studies adopted random sampling [15-17,20-23,25,27,29-32, 36,38,40-42,44,48,50-53,56,57,60-62,64,65,67,69,71,72], 12 used K-fold cross-validation [14,24,34,37,39,46,49,58,59,66, 70,73], and 5 applied external validation [28,33,45,54, 63]. In total, 9 studies compared their results with the screening results of clinicians [22,25,28,45-47,60,62,65]. In terms of model construction, 32 built DL models [18,21, 23-31,33,37-42,44-47,52-56,59-62,66], and 28 constructed ML models [14-17,19,20,22,32,34-36,43,48-51,57,58,63-65, 67-73] (Multimedia Appendix 2).

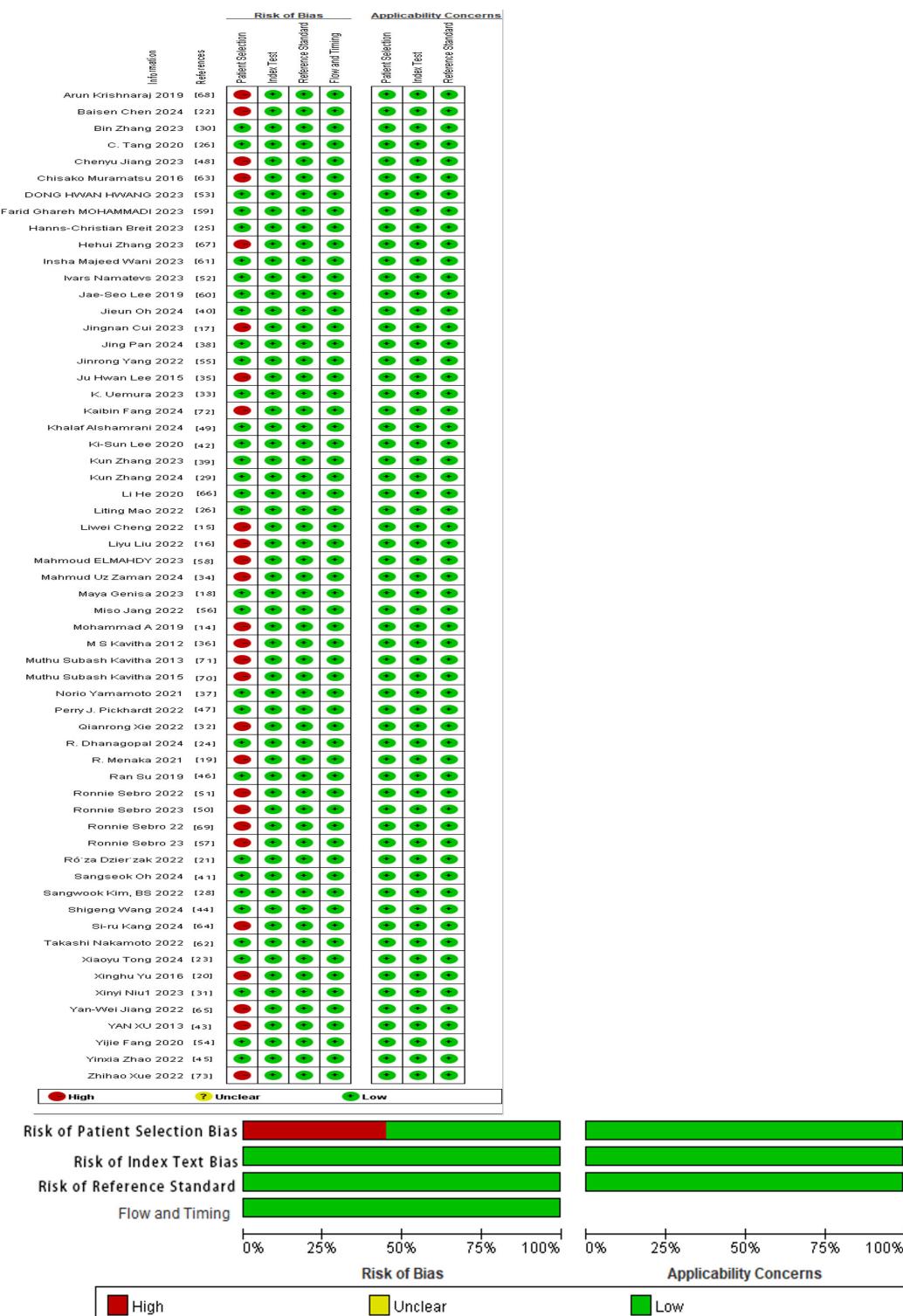
## Risk of Bias in Studies

In all eligible studies, consecutive cases were included. Although most studies were case-control studies, 32 developed DL models, with variables derived from medical images. Therefore, these studies demonstrated a low risk of bias in case selection. In total, 28 studies applied ML models, where the process of variable generation might be influenced by the case-control study design, thereby leading

to a higher risk of bias. Since this research is a meta-analysis of ML, whether or not the reference standards for OP diagnosis are known does not affect the results. Additionally, the criteria for determining positive results were pre-established, indicating that the trials under evaluation posed a low risk of bias. The implementation of a reference standard for

OP diagnosis was considered reasonable, thus introducing a low risk of bias. Furthermore, there was a proper time interval between the trial and reference standard, and all patients in a given study followed the same diagnosis rules, with no cases omitted. Therefore, there was a low risk of bias in clinical applicability [14-73] (Figure 2 and [Multimedia Appendix 3](#)).

**Figure 2.** Risk of bias plot [14-73].



## Meta-Analysis

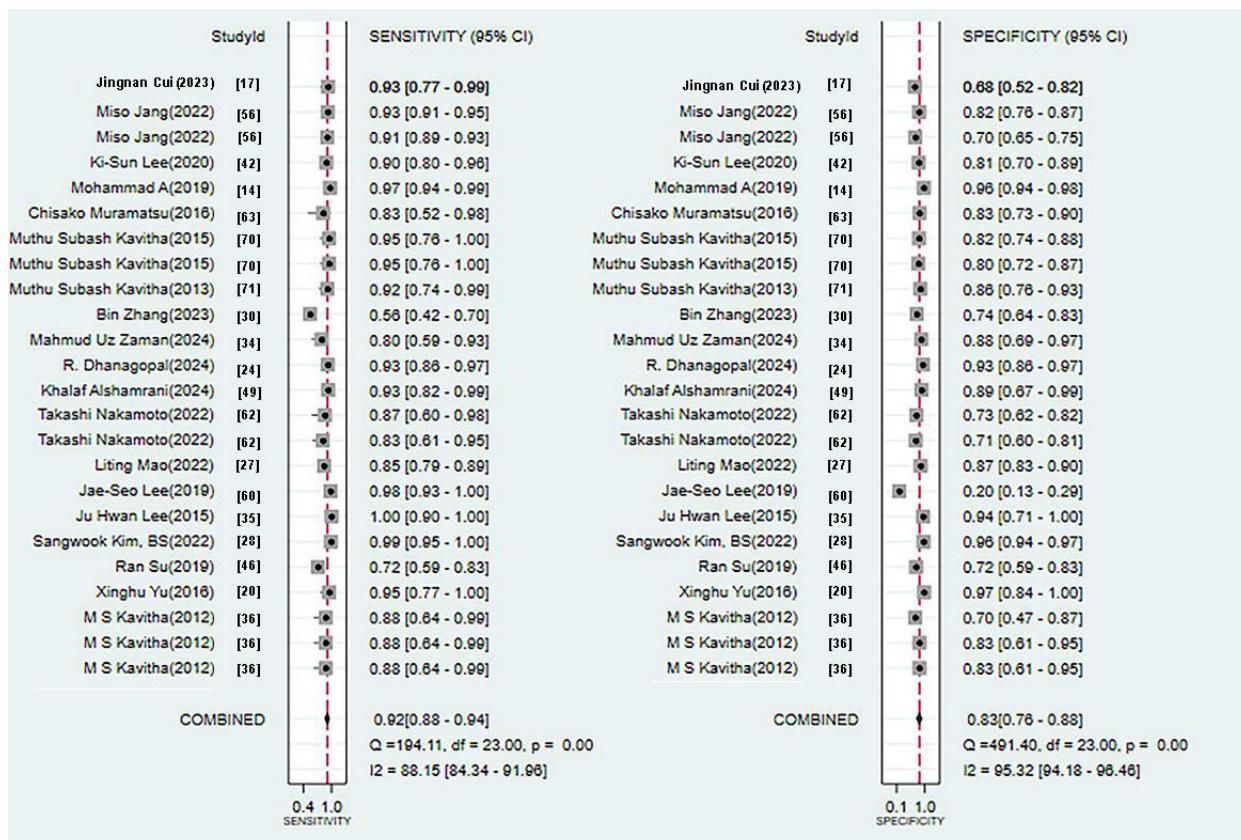
### ML Based on X-Ray

#### Synthesized Results

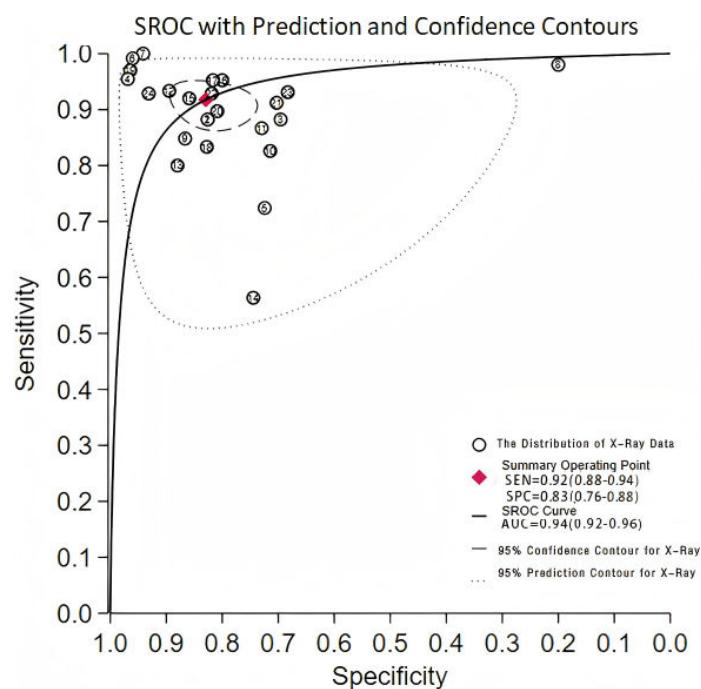
This validation set comprised 24 diagnostic 4-fold tables, which were used to verify the ML models based on x-rays for OP diagnosis. The results were summarized through the bivariate mixed-effects model. The pooled SEN, SPC, PLR, NLR, DOR, and SROC curves were 0.92 (95% CI 0.88-0.94),

0.83 (95% CI 0.76-0.88), 5.4 (95% CI 3.8-7.6), 0.10 (95% CI 0.07-0.15), 54 (95% CI 28-105), and 0.94 (95% CI 0.92-0.96), respectively (Figures 3 and 4) [14,17,20,24,27,28, 30,34-37,42,46,49,55,59-63,70,71]. There was no discernible publication bias in the studies according to Deeks' funnel plot (Figure 5). In the included study participants, approximately 48.44% (n=6429) had OP. Assuming this as the prior probability, if the ML prediction result was OP, the actual probability of OP was .83. If the ML prediction result was non-OP, the actual probability of non-OP was .92 (Figure 6).

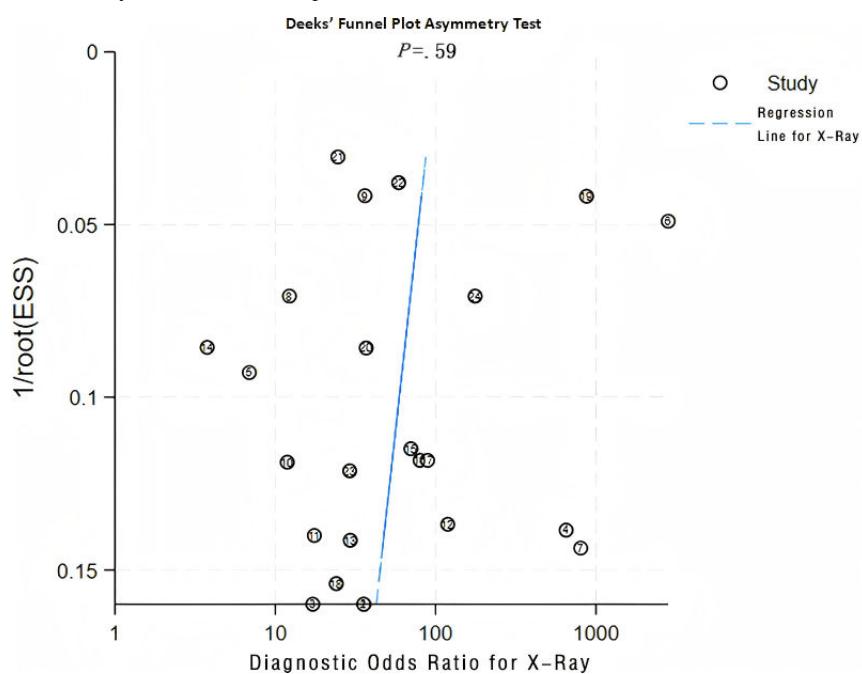
**Figure 3.** Forest plot of sensitivity and specificity for x-ray-based machine learning for diagnosing osteoporosis [14,17,20,24,27,28,30,34-36,42,46, 49,56,60,62,63,70,71].

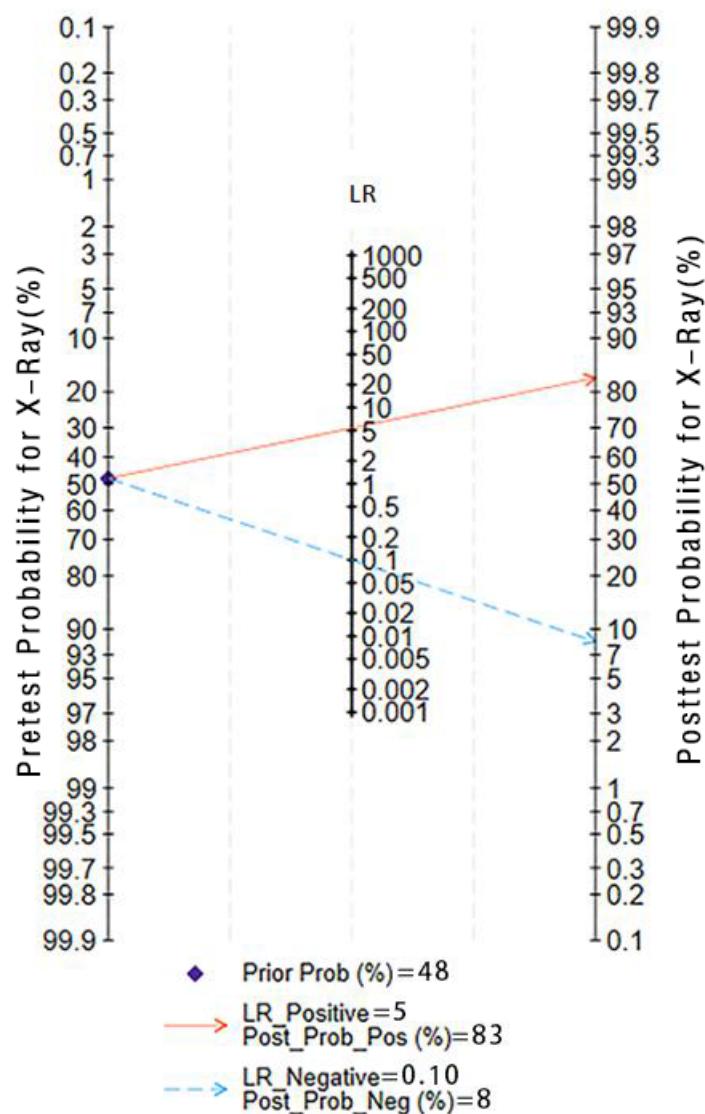


**Figure 4.** 95% Confidence contour for x-ray. AUC: Area under the curve; SEN: Sensitivity; SPC: Specificity; SROC: Summary receiver operating characteristic.



**Figure 5.** Diagnostic odds ratio for x-ray. ESS: Effective sample size.



**Figure 6.** Pretest probability for x-ray. LR:Likelihood ratio.

### Subgroup Analysis: Types of ML

#### Deep Learning

The validation set included 9 diagnostic 4-fold tables to assess the performance of DL models based on x-ray images for OP diagnosis. The results summarized from the bivariate mixed-effects model showed that SEN, SPC, PLR, NLR, DOR, and the SROC curve were 0.90 (95% CI 0.79-0.95), 0.79 (95% CI 0.62-0.89), 4.2 (95% CI 2.2-8.0), 0.13 (95% CI 0.06-0.29), 32 (95% CI 9-107), and 0.92 (95% CI 0.89-0.94), respectively (Figures S1 and S2 in [Multimedia Appendix 4](#)). Deeks' funnel plot revealed no marked publication bias (Figure S3 in [Multimedia Appendix 4](#)). In the encompassed studies, approximately 30% (n=2556) of the participants had OP. Therefore, assuming this as the prior probability, if the result from ML indicated OP, the actual probability of OP was .64. If the ML result indicated non-OP, the actual probability of non-OP was .95 (Figure S4 in [Multimedia Appendix 4](#)).

#### Traditional ML

The validation set encompassed 15 diagnostic 4-fold tables for validating the traditional ML models based on x-ray imaging for OP diagnosis. The bivariate mixed-effects model was leveraged. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.93 (95% CI 0.92-0.95), 0.85 (95% CI 0.79-0.89), 6.0 (95% CI 4.3-8.5), 0.08 (95% CI 0.06-0.10), 78 (95% CI 44-139), and 0.96 (95% CI 0.94-0.97), respectively (Figures S5 and S6 in [Multimedia Appendix 4](#)). Deeks' funnel plot showed no significant publication bias in studies (Figure S7 in [Multimedia Appendix 4](#)). Approximately 61% (n=3863) of the participants had OP. Therefore, assuming this as the prior probability, if the result from ML indicated OP, the actual probability of OP was .90. If the ML result indicated non-OP, the actual probability of non-OP was .81 (Figure S8 in [Multimedia Appendix 4](#)).

## Generation Method of the Validation Set

### K-Fold Cross-Validation

Among models constructed through x-ray for OP diagnosis, 7 diagnostic 4-fold tables used the K-fold cross-validation to generate the validation set. The results summarized by the bivariate mixed-effects model demonstrated that the SEN, SPC, PLR, NLR, DOR, and SROC curve were 0.90 (95% CI 0.83-0.95), 0.87 (95% CI 0.79-0.93), 7.2 (95% CI 4.0-12.7), 0.11 (95% CI 0.06-0.21), 64 (95% CI 20-204), and 0.95 (95% CI 0.93-0.96), respectively (Figures S9 and S10 in [Multimedia Appendix 4](#)). Deeks' funnel plot did not exhibit significant publication bias (Figure S11 in [Multimedia Appendix 4](#)). Among the participants in our included studies, approximately 42% (n=1287) had OP. Therefore, assuming this as the prior probability, if the ML result indicated OP, the actual probability of OP was .84. If the ML result indicated non-OP, the actual probability of non-OP was .92 (Figure S12 in [Multimedia Appendix 4](#)).

### Random Sampling

In total, 14 diagnostic 4-fold tables used the random sampling method to generate the validation set. The results summarized by the bivariate mixed-effects model showed that the pooled SEN, SPC, PLR, NLR, DOR, and SROC curve were 0.90 (95% CI 0.84-0.93), 0.76 (95% CI 0.67-0.84), 3.8 (95% CI 2.7-5.4), 0.14 (95% CI 0.09-0.20), 28 (95% CI 16-48), and 0.91 (95% CI 0.88-0.93), respectively (Figures S13 and S14 in [Multimedia Appendix 4](#)). Significant publication bias was not noted in Deeks' funnel plot (Figure S15 in [Multimedia Appendix 4](#)). Among the participants in our included studies, approximately 60% (n=4049) had OP. Therefore, assuming this as the prior probability, if the ML result indicated OP, the actual probability of OP was .85. If the ML result showed non-OP, the actual probability of non-OP was .83 (Figure S16 in [Multimedia Appendix 4](#)).

### Examination Parts

#### Limbs

In the OP diagnosis models constructed based on x-rays, 4 diagnostic 4-fold tables focused on the limb bones. The results summarized by the bivariate mixed-effects model showed a SEN of 0.97 (95% CI 0.83-0.99), SPC of 0.90 (95% CI 0.75-0.96), PLR of 9.6 (95% CI 3.5-25.9), NLR of 0.03 (95% CI 0.01-0.22), DOR of 277 (95% CI 20-3783), and the SROC curve of 0.98 (95% CI 0.96-0.99; Figures S17 and S18 in [Multimedia Appendix 4](#)). Deeks' funnel plot did not demonstrate significant publication bias (Figure S19 in [Multimedia Appendix 4](#)). In the included study participants, the proportion of OP cases was approximately 19% (n=1114). Assuming this as the prior probability, if the ML result indicated OP, the actual probability of OP was .69. If the ML result indicated non-OP, the actual probability of non-OP was <.001 (Figure S20 in [Multimedia Appendix 4](#)).

#### Mandible

In total, 6 diagnostic 4-fold tables focused on the mandible. The bivariate mixed-effects model was used. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.94 (95% CI 0.89-0.97), 0.80 (95% CI 0.56-0.93), 4.8 (95% CI 1.9-12.1), 0.07 (95% CI 0.04-0.14), 69 (95% CI 20-241), and 0.96 (95% CI 0.94-0.97), respectively (Figures S21 and S22 in [Multimedia Appendix 4](#)). Deeks' funnel plot indicated no significant publication bias (Figure S23 in [Multimedia Appendix 4](#)). In all included study participants, the proportion of OP cases was approximately 42% (n=1153). Assuming this as the prior probability, if the ML result indicated OP, the actual probability of OP was .78. If the ML result indicated non-OP, the actual probability of non-OP was .95 (Figure S24 in [Multimedia Appendix 4](#)).

#### Lumbar Vertebrae

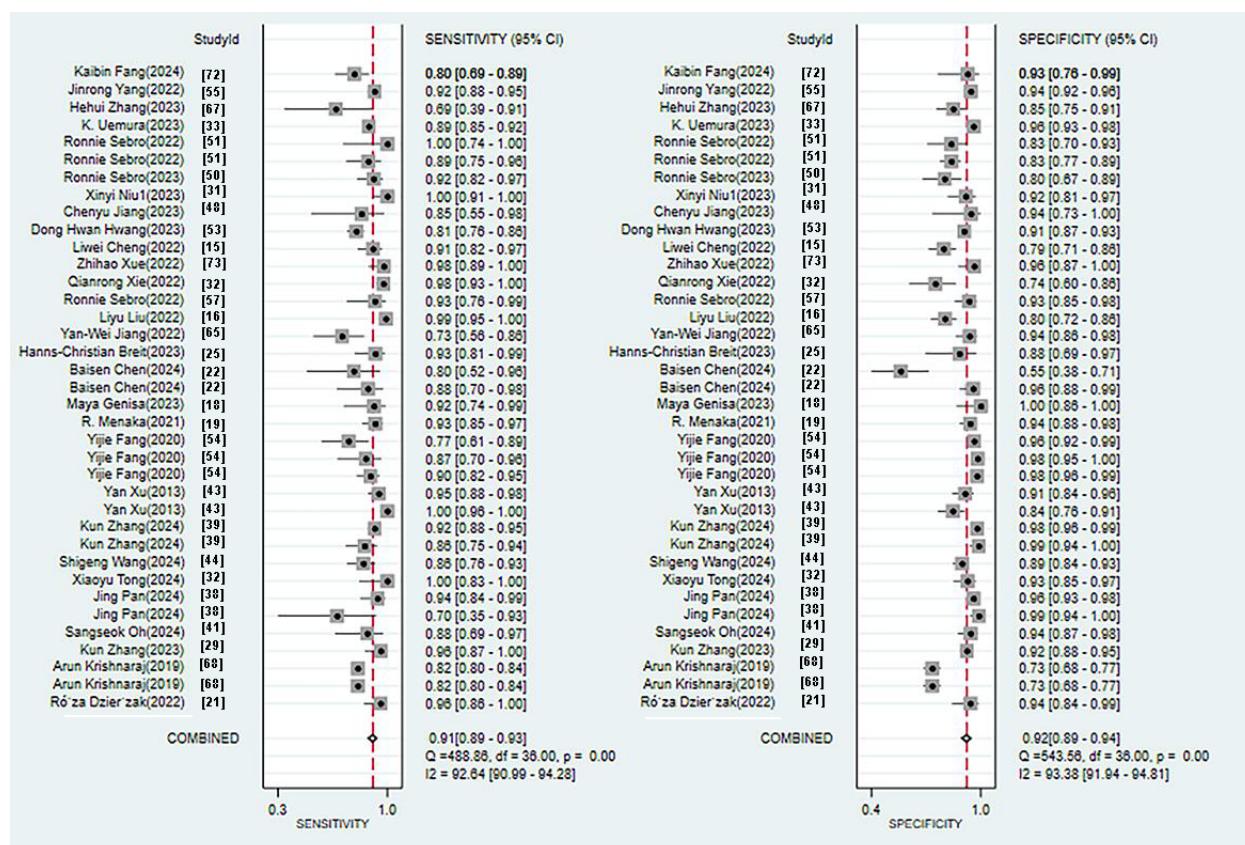
In total, 8 diagnostic 4-fold tables focused on the lumbar vertebrae. The bivariate mixed-effects model was used to summarize data. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.87 (95% CI 0.77-0.93), 0.82 (95% CI 0.75-0.87), 4.8 (95% CI 3.4-6.7), 0.16 (95% CI 0.08-0.30), 31 (95% CI 12-77), and 0.90 (95% CI 0.87-0.92), respectively (Figures S25 and S26 in [Multimedia Appendix 4](#)). Significant publication bias was not found in Deeks' funnel plot (Figure S27 in [Multimedia Appendix 4](#)). In the included study participants, the proportion of OP cases was approximately 32% (n=1281). Assuming this as the prior probability, if the ML result indicated OP, the actual probability of having OP was .69. If the ML result indicated non-OP, the actual probability of non-OP was .93 (Figure S28 in [Multimedia Appendix 4](#)).

### ML Based on CT

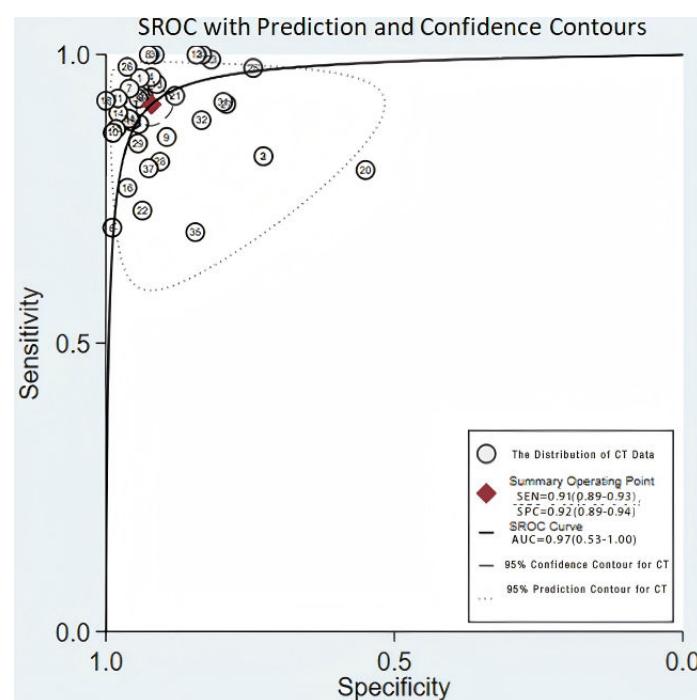
#### Synthesized Results

The validation set consisted of 24 diagnostic 4-fold tables for validating CT-based ML models for diagnosing OP. The bivariate mixed-effects model was used to pool data. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.91 (95% CI 0.89-0.93), 0.92 (95% CI 0.89-0.94), 11.6 (95% CI 8.5-9.7), 0.09 (95% CI 0.07-0.12), 123 (95% CI 80-90), and 0.97 (95% CI 0.53-1.00), respectively (Figures 7 and 8) [15,16,18,19,21-23,25,26,29,31-33,38-41,43-45,47,48,50-54,56-58,65,67-69,72,73]. According to Deeks' funnel plot, there was no significant publication bias (Figure 9). Among the included research participants, the proportion of individuals with OP was approximately 50% (n=10,995). Therefore, assuming this as the prior probability, if the ML models predicted OP, the actual probability of OP was .92. If the ML models predicted no OP, the actual probability of non-OP was .91 (Figure 10).

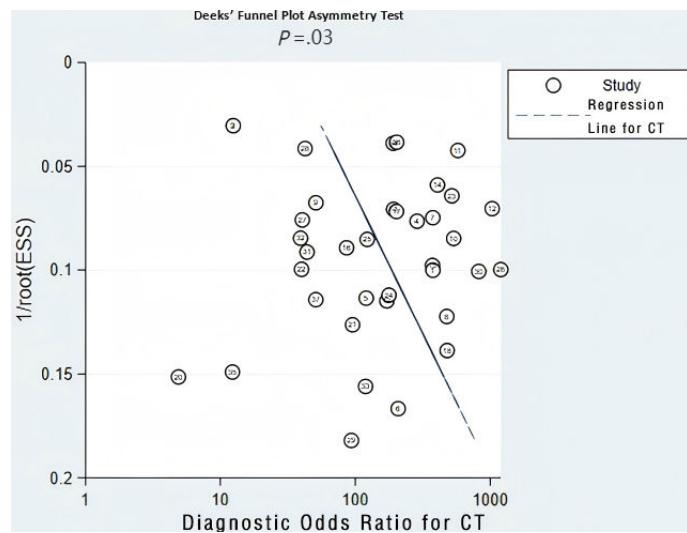
**Figure 7.** Forestplot of sensitivity and specificity for computed tomography-based machine learning for diagnosing osteoporosis [15,16,18,19,21-23, 25,26,29,31-33,38-41,43-45,47,48,50-54,56-58,65,67-69,72,73].



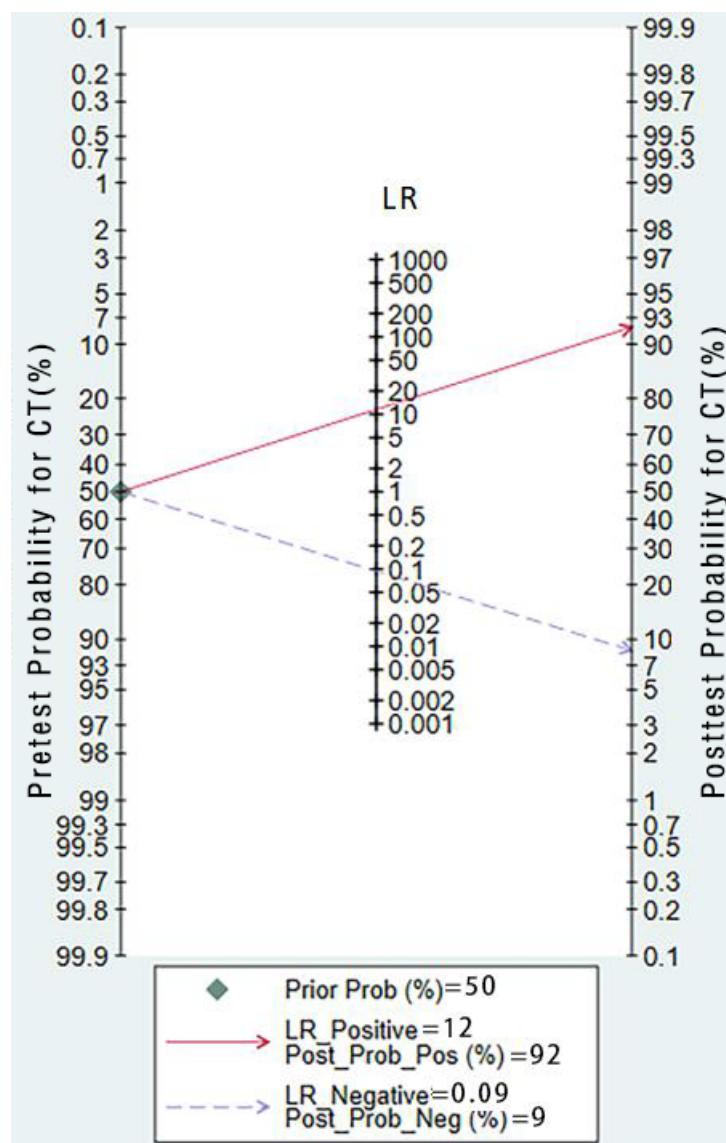
**Figure 8.** 95% Confidence contour for CT.AUC:Area under the curve;CT:Computed tomography;SEN: Sensitivity; SPC: Specificity;SROC: Summary receiver operating characteristic.



**Figure 9.** Diagnostic odds ratio for CT .CT: Computed tomography;ESS:Effective sample size.



**Figure 10.** Pretest probability for CT. CT:Computed tomography;LR:Likelihood ratio.



## Subgroup Analysis: Types of ML

### Deep Learning

In the validation set, there were 15 diagnostic 4-fold tables for validating the CT-based DL models for diagnosing OP. The bivariate mixed-effects model was used. The pooled SEN, SPC, PLR, NLR, DOR, and SROC curve were 0.91 (95% CI 0.88-0.94), 0.94 (95% CI 0.92-0.96), 16.3 (95% CI 11.9-22.3), 0.09 (95% CI 0.06-0.13), 178 (95% CI 106-299), and 0.98 (95% CI 0.96-0.99), respectively (Figures S29 and S30 in [Multimedia Appendix 4](#)). Deeks' funnel plot exhibited no marked publication bias (Figure S31 in [Multimedia Appendix 4](#)). Among the research participants included, the proportion of individuals with OP was approximately 32% (n=3197). Therefore, assuming this as the prior probability, if the ML models predicted OP, the actual probability of OP was .88. If the ML models predicted no OP, the actual probability of non-OP was .96 (Figure S32 in [Multimedia Appendix 4](#)).

### Traditional ML

In the validation set, there were 15 diagnostic 4-fold tables for validating traditional ML models based on CT for diagnosing OP. The bivariate mixed-effects model was used. The pooled SEN, SPC, PLR, NLR, DOR, and SROC curve were 0.92 (95% CI 0.88-0.95), 0.85 (95% CI 0.77-0.90), 6.1 (95% CI 4.0-9.4), 0.09 (95% CI 0.06-0.15), 67 (95% CI 35-128), and 0.95 (95% CI 0.93-0.97), respectively (Figures S33 and S34 in [Multimedia Appendix 4](#)). Deeks' funnel plot did not show notable publication bias (Figure S35 in [Multimedia Appendix 4](#)). Among the research participants, the proportion of individuals with OP was approximately 60% (n=6486). Therefore, assuming this as the prior probability, if the ML models predicted OP, the actual probability of OP was .90. If the ML models predicted no OP, the actual probability of non-OP was .88 (Figure S36 in [Multimedia Appendix 4](#)).

### Validation Set Generation Method

#### External Validation

In the OP diagnosis models constructed based on CT, validation sets for 5 diagnostic 4-fold tables were generated through external validation. The bivariate mixed-effects model was leveraged to pool data. The pooled SEN, SPC, PLR, NLR, DOR, and SROC curve were 0.88 (95% CI 0.85-0.91), 0.97 (95% CI 0.96-0.98), 28.4 (95% CI 20.4-39.7), 0.12 (95% CI 0.10-0.16), 229 (95% CI 148-355), and 0.98 (95% CI 0.96-0.99), respectively (Figures S37 and S38 in [Multimedia Appendix 4](#)). Deeks' funnel plot indicated no discernible publication bias (Figure S39 in [Multimedia Appendix 4](#)). Among the included study participants, approximately 31% (n=1590) had OP. Assuming this as the prior probability, if ML predicted OP, the actual probability of OP was .93. Conversely, if ML predicted non-OP, the actual probability of non-OP was .95 (Figure S40 in [Multimedia Appendix 4](#)).

## Random Sampling

Validation sets for 24 diagnostic 4-fold tables were generated using the random sampling method. The bivariate mixed-effects model was leveraged. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.91 (95% CI 0.87-0.94), 0.90 (95% CI 0.85-0.94), 9.4 (95% CI 6.0-14.9), 0.10 (95% CI 0.07-0.14), 96 (95% CI 57-161), and 0.96 (95% CI 0.94-0.97), respectively (Figures S41 and S42 in [Multimedia Appendix 4](#)). Deeks' funnel plot presented no significant publication bias (Figure S43 in [Multimedia Appendix 4](#)). Among the included study participants, approximately 36% (n=4175) had OP. Given this as the prior probability, when the ML models predicted OP, the actual probability of OP was .84. On the other hand, when the ML models predicted non-OP, the actual probability of non-OP was .95 (Figure S44 in [Multimedia Appendix 4](#)).

## Examination Parts

### Hip Joint

In the OP diagnostic models constructed based on CT, 6 diagnostic 4-fold tables focused on the hip joint. The bivariate mixed-effects model was used. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.87 (95% CI 0.83-0.90), 0.92 (95% CI 0.81-0.96), 10.4 (95% CI 4.4-24.7), 0.14 (95% CI 0.10-0.19), 76 (95% CI 24-239), and 0.92 (95% CI 0.90-0.94), respectively (Figures S45 and S46 in [Multimedia Appendix 4](#)). Deeks' funnel plot did not show any marked publication bias (Figure S47 in [Multimedia Appendix 4](#)). Among the included study participants, approximately 69% (n=2719) had OP. Assuming this as the prior probability, if ML predicted OP, the actual probability of OP was .96. If the models predicted non-OP, the actual probability of OP was .77 (Figure S48 in [Multimedia Appendix 4](#)).

### Thoracic Vertebrae

In total, 9 diagnostic 4-fold tables focused on the thoracic vertebrae. The bivariate mixed-effects model was leveraged to pool data. The pooled SEN, SPC, PLR, NLR, DOR, and SROC were 0.91 (95% CI 0.86-0.94), 0.94 (95% CI 0.92-0.95), 14.4 (95% CI 10.7-19.3), 0.10 (95% CI 0.06-0.15), 150 (95% CI 75-300), and 0.97 (95% CI 0.95-0.98), respectively (Figures S49 and S50 in [Multimedia Appendix 4](#)). Significant publication bias was not observed in Deeks' funnel plot (Figure S51 in [Multimedia Appendix 4](#)). Among the encompassed study participants, approximately 29% (n=2523) had OP. Assuming this as the prior probability, if ML predicted OP, the actual probability of OP was .85. If ML predicted non-OP, the actual probability of non-OP was .96 (Figure S52 in [Multimedia Appendix 4](#)).

### Lumbar Vertebrae

For OP diagnostic models using the lumbar vertebrae as the target part, 26 diagnostic 4-fold tables were analyzed. The bivariate mixed-effects model yielded a SEN of 0.91 (95% CI 0.87-0.94), SPC of 0.92 (95% CI 0.86-0.95), PLR of 10.7 (95% CI 6.7-17.2), NLR of 0.10 (95% CI 0.07-0.14), DOR of 110 (95% CI 63-191), and SROC curve of 0.96

(95% CI 0.94-0.98; Figures S53 and S54 in [Multimedia Appendix 4](#)). Deeks' funnel plot did not reflect discernible publication bias (Figure S55 in [Multimedia Appendix 4](#)). Among the encompassed study participants, approximately 42% (n=7327) had OP. Assuming this as the prior probability, if ML predicted OP, the probability of actual OP was .89. Conversely, if ML predicted non-OP, the actual likelihood of non-OP was .93 (Figure S56 in [Multimedia Appendix 4](#)).

## ML Based on MRI

Only 3 studies have constructed diagnostic models for OP based on MRI [45,64,66], all of which used the lumbar vertebrae as the examination part. Due to the limited number of studies of this type and the substantial heterogeneity noted in the meta-analysis, the conclusions drawn lack sufficient reference significance. Therefore, this study presents only a narrative analysis of this part.

Among these, 2 studies used traditional ML models, while 1 used a DL model. The SEN of these models was 0.857, 0.872, and 0.892, and the SPC was 0.944, 0.688, and 0.892, respectively. The validation strategies used in these studies included external validation, K-fold cross-validation, and random sampling.

## Discussion

### Main Findings of This Study

Medical imaging is an indispensable tool in the diagnosis, treatment, and management of OP. Conventional imaging methods such as x-ray, CT, and MRI are pivotal clinically.

X-ray imaging enables clinicians to visually assess reductions in vertebral height, cortical bone thickness, and morphological changes in appendicular and mandibular bones, thus screening OP. However, as DXA is updated and improved, clinicians can more accurately know bone mineral density and structural parameters of the lumbar vertebrae and hip, thereby facilitating the diagnosis of OP. The World Health Organization has designated DXA as the gold standard for determining bone mineral density and diagnosing postmenopausal OP [74,75]. However, in low-resource environments and economically underdeveloped regions, the clinical application of DXA is limited due to factors such as insufficient medical knowledge and constrained health care infrastructure. In contrast, AI tools have the potential to maximize the extraction of clinically relevant information from various medical images, thereby enabling the early identification of the population with OP or low bone mass. This significantly supports the early prevention, diagnosis, and management of the disease.

### Advantages of Different Imaging Modalities in the Diagnosis of OP

This shift has diminished the application of x-rays in quantitative analysis for OP. Nevertheless, ML and DL models have improved the diagnostic performance of x-ray imaging, providing significant impetus for its broader clinical application. CT, with its high resolution, enables clinicians

to observe cortical and trabecular bone integrity, offering distinct advantages in evaluating spinal OP and changes in trabecular bone volume ratios in the hip [76]. Conventional CT generates images by measuring differences in the linear attenuation coefficients of x-ray beams as they pass through various biological tissues. However, when tissues possess similar densities, such as calcium and bone, conventional CT often yields comparable Hounsfield unit values due to the use of a single x-ray energy spectrum, limiting its ability to differentiate between such tissues. In contrast, spectral CT imaging, which is based on tissue-specific photoelectric effect weighting, offers enhanced resolution in distinguishing fine bone microarchitecture. This technological advancement holds significant potential for improving the diagnostic accuracy of OP. MRI is highly efficient in assessing bone microarchitecture [77]. However, MRI is not the first choice to detect OP because of its high cost, extended scan times, and obstacles faced by patients with metallic implants or claustrophobia. Our database search corroborated that most studies have focused on x-ray and CT imaging, while comparatively fewer have investigated MRI. Nevertheless, existing evidence supports the robust diagnostic performance of ML models based on imaging data. For example, the pooled SEN and SPC of ML models based on x-ray for OP diagnosis were 0.92 (95% CI 0.88-0.94) and 0.83 (95% CI 0.76-0.88), respectively. Similarly, ML models developed via CT achieved SEN and SPC of 0.91 (95% CI 0.89-0.93) and 0.92 (95% CI 0.89-0.94), respectively. These findings demonstrate the high accuracy of x-ray and CT in OP diagnosis. In addition, quantitative ultrasound is another commonly used modality for OP detection. Quantitative ultrasound relies on 2 primary parameters: speed of sound and broadband ultrasound attenuation, which assess the ability of ultrasound waves to propagate through bone both horizontally and longitudinally [78]. In summary, diverse imaging modalities and bone types provide flexible and enriched diagnostic options for OP. Furthermore, this variety brings ample opportunities for the development of advanced ML models tailored to different imaging techniques.

### Status Quo of Research on ML

With advances in computer science, numerous researchers have sought to use these techniques in the prevention and treatment of OP. Compared with clinicians, who visually observe positive imaging features, AI-assisted tools significantly improve the efficiency and accuracy of diagnosing OP [46,60]. In addition, Yang et al [79] developed an ML-based predictive model using data from surveys on risk factors for OP, which is highly prospective for early screening and treating OP in the Hong Kong population. Similarly, ML models based on community health examinations and serum bone turnover markers have demonstrated a high area under the receiver operating characteristic curve,  $F_1$ -scores, and accuracy [80,81]. These findings highlight the efficiency of ML in the diagnosis and management of OP.

### Mechanism of Image-Based ML

Image-based ML can broadly be categorized into traditional ML and DL. Traditional ML involves dividing data into

a training set for model development and a test set for model validation. Through processes such as image segmentation, texture extraction, and feature selection, traditional ML models are constructed for predicting outcome events. However, the process of texture feature extraction and selection carries a significant risk of data loss. In contrast, DL incorporates feature extraction directly into the training process, thereby maximizing the retention of meaningful information within the image data. Convolutional neural networks, as a representative DL approach, can simultaneously extract and select features across multiple hidden layers to accomplish classification tasks. Moreover, DL-based models can correct image blurring in panoramic x-rays caused by patient mispositioning and mitigate the impact of metal artifacts in CT images on feature extraction [82-84]. This study further demonstrates that ML models based on x-ray and CT outperform traditional ML models, suggesting that DL is more accurate than traditional ML approaches. Image analysis using DL can leverage AI to develop more efficient and user-friendly image interpretation tools, providing valuable insights into the development of medical imaging software.

### ***The Impact of Validation Set Generation Methods on ML Performance***

Validation methods are critical metrics for assessing the performance of ML models. These methods can be categorized into external validation and internal validation. Internal validation can be further subdivided into random sampling, leave-one-out validation, and K-fold cross-validation. External validation, which can accurately reflect the clinical applicability of ML, is widely preferred by researchers. In contrast, internal validation typically generates validation sets via random methods, which inherently carries a risk of similarity in features and distribution trends between the validation and training sets. This issue is prominent in image-based studies, where the application of ML is restricted in medical research due to the high similarity in images and parameters between internal validation sets. Although external validation offers a superior means of assessing model

performance, conducting such validation requires access to independent research cohorts and often entails consideration of factors such as periods, geographical regions, populations, and health care institutions. These requirements inevitably lead to substantial increases in both the time and financial costs of research. This perspective provides an objective explanation for the limited external validation in this study.

### ***Advantages and Limitations***

This study is the first to summarize the evidence of the application of ML based on various imaging modalities in the diagnosis of OP. This study provides theoretical support for the subsequent development of clinical scoring systems and medical software. However, our research has the following limitations: first, despite a substantial number of included studies, only a small number of studies on MRI were encompassed in view of the practicability in clinical work. Therefore, in future research, our emphasis will be put on meta-analyses involving MRI studies, aiming to evaluate the utility of ML in the diagnosis of OP through medical imaging. As a result, only a narrative review was performed, without a direct evaluation of its diagnostic performance. Most of the included studies rely primarily on internal validation, with insufficient external validation, which imposes certain limitations on the interpretability and generalizability of our findings. This study encompassed only English publications, with the majority of research originating from countries where AI is more widely applied. In addition, the external validation conducted in this study was limited, constituting an objective constraint that may have influenced the outcomes of the meta-analysis. Future studies will endeavor to comprehensively incorporate globally available literature to enhance the authority and generalizability of the conclusions.

### ***Conclusions***

Image-based ML, particularly DL based on x-ray and CT images, is highly accurate in the diagnosis of OP. Future focus should be placed on developing AI-based software to expand its clinical applicability and enhance diagnostic precision.

### **Acknowledgments**

This work was supported by the National Natural Science Foundation of China (82274551) and the Guangdong Provincial Fundamental and Applied Basic Research Fund Program (2023B1515230001). All authors of this study hereby solemnly declare that no forms of artificial intelligence tools were used at any stage of the research process to ensure the originality and independence of this manuscript.

### **Data Availability**

The datasets used and analyzed during this study are available from the corresponding author upon reasonable request.

### **Authors' Contributions**

RZ was responsible for the formal analysis, data curation, and original manuscript writing of all the experiments. HY was responsible for software operation and data management. Y Li and XL were responsible for the formal analysis. ZY was in charge of the methodology. Y Lin and JH were responsible for the methodology and data validation. LW was in charge of project management. HH was responsible for project design, project management, and fund acquisition. All authors reviewed the final manuscript.

### **Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Literature search strategy.

[[DOCX File \(Microsoft Word File\), 25 KB](#)-[Multimedia Appendix 1](#)]

**Multimedia Appendix 2**

Supplementary data sheet.

[[XLSX File \(Microsoft Excel File\), 20 KB](#)-[Multimedia Appendix 2](#)]

**Multimedia Appendix 3**

Quality Assessment of Diagnostic Accuracy Studies assessment process for included studies.

[[XLSX File \(Microsoft Excel File\), 16 KB](#)-[Multimedia Appendix 3](#)]

**Multimedia Appendix 4**

Supplementary materials.

[[DOC File \(Microsoft Word File\), 56032 KB](#)-[Multimedia Appendix 4](#)]

**Checklist 1**

PRISMA checklist.

[[PDF File \(Adobe File\), 99 KB](#)-[Checklist 1](#)]

**References**

1. Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med*. Jun 1993;94(6):646-650. [doi: [10.1016/0002-9343\(93\)90218-E](https://doi.org/10.1016/0002-9343(93)90218-E)]
2. Wang Y, Tao Y, Hyman ME, Li J, Chen Y. Osteoporosis in china. *Osteoporos Int*. Oct 2009;20(10):1651-1662. [doi: [10.1007/s00198-009-0925-y](https://doi.org/10.1007/s00198-009-0925-y)] [Medline: [19415374](#)]
3. Clynes MA, Harvey NC, Curtis EM, Fuggle NR, Dennison EM, Cooper C. The epidemiology of osteoporosis. *Br Med Bull*. May 15, 2020;133(1):105-117. [doi: [10.1093/bmb/ldaa005](https://doi.org/10.1093/bmb/ldaa005)] [Medline: [32282039](#)]
4. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. *World Health Organ Tech Rep Ser*. 1994;843(1-129):1-129. [Medline: [7941614](#)]
5. Kanis JA, Cooper C, Rizzoli R, Reginster JY, on behalf of the 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). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int*. Jan 18, 2019;30(1):3-44. [doi: [10.1007/s00198-018-4704-5](https://doi.org/10.1007/s00198-018-4704-5)]
6. Erickson BJ. Basic artificial intelligence techniques: machine learning and deep learning. *Radiol Clin North Am*. Nov 2021;59(6):933-940. [doi: [10.1016/j.rcl.2021.06.004](https://doi.org/10.1016/j.rcl.2021.06.004)] [Medline: [34689878](#)]
7. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. *Adv Exp Med Biol*. 2020;1213(3-21):3-21. [doi: [10.1007/978-3-030-33128-3\\_1](https://doi.org/10.1007/978-3-030-33128-3_1)] [Medline: [32030660](#)]
8. Oliveira JS, Franco FO, Revers MC, et al. Computer-aided autism diagnosis based on visual attention models using eye tracking. *Sci Rep*. May 12, 2021;11(1):10131. [doi: [10.1038/s41598-021-89023-8](https://doi.org/10.1038/s41598-021-89023-8)] [Medline: [33980874](#)]
9. Khan M, Shah PM, Khan IA, et al. IoMT-enabled computer-aided diagnosis of pulmonary embolism from computed tomography scans using deep learning. *Sensors (Basel)*. Jan 28, 2023;23(3):1471. [doi: [10.3390/s23031471](https://doi.org/10.3390/s23031471)] [Medline: [36772510](#)]
10. Loizidou K, Skouroumouni G, Nikolaou C, Pitris C. A review of computer-aided breast cancer diagnosis using sequential mammograms. *Tomography*. Dec 6, 2022;8(6):2874-2892. [doi: [10.3390/tomography8060241](https://doi.org/10.3390/tomography8060241)] [Medline: [36548533](#)]
11. Ceranka J, Wuts J, Chiabai O, Lecouvet F, Vandemeulebroucke J. Computer-aided diagnosis of skeletal metastases in multi-parametric whole-body MRI. *Comput Methods Programs Biomed*. Dec 2023;242:107811. [doi: [10.1016/j.cmpb.2023.107811](https://doi.org/10.1016/j.cmpb.2023.107811)] [Medline: [37742486](#)]
12. Hong N, Cho SW, Shin S, et al. Deep-learning-based detection of vertebral fracture and osteoporosis using lateral spine X-ray radiography. *J Bone Miner Res*. Dec 1, 2020;38(6):887-895. [doi: [10.1002/jbmr.4814](https://doi.org/10.1002/jbmr.4814)]
13. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. Oct 18, 2011;155(8):529-536. [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](#)]
14. Alzubaidi MA, Otoom M. A comprehensive study on feature types for osteoporosis classification in dental panoramic radiographs. *Comput Methods Programs Biomed*. May 2020;188(105301):105301. [doi: [10.1016/j.cmpb.2019.105301](https://doi.org/10.1016/j.cmpb.2019.105301)] [Medline: [31911333](#)]

15. Cheng L, Cai F, Xu M, Liu P, Liao J, Zong S. A diagnostic approach integrated multimodal radiomics with machine learning models based on lumbar spine CT and X-ray for osteoporosis. *J Bone Miner Metab*. Nov 2023;41(6):877-889. [doi: [10.1007/s00774-023-01469-0](https://doi.org/10.1007/s00774-023-01469-0)] [Medline: [37898574](#)]
16. Liu L, Si M, Ma H, et al. A hierarchical opportunistic screening model for osteoporosis using machine learning applied to clinical data and CT images. *BMC Bioinformatics*. Feb 10, 2022;23(1):63. [doi: [10.1186/s12859-022-04596-z](https://doi.org/10.1186/s12859-022-04596-z)] [Medline: [35144529](#)]
17. Cui J, Liu CL, Jennane R, Ai S, Dai K, Tsai TY. A highly generalized classifier for osteoporosis radiography based on multiscale fractal, lacunarity, and entropy distributions. *Front Bioeng Biotechnol*. 2023;11:1054991. [doi: [10.3389/fbioe.2023.1054991](https://doi.org/10.3389/fbioe.2023.1054991)] [Medline: [37274169](#)]
18. Genisa M, Abdullah JY, Yusoff BM, Arief EM, Hermana M, Utomo CP. Adopting signal processing technique for osteoporosis detection based on CT scan image. *Appl Sci (Basel)*. 2023;13(8):5094. [doi: [10.3390/app13085094](https://doi.org/10.3390/app13085094)]
19. Menaka R, Ramesh R, Dhanagopal R. Aggregation of region-based and boundary-based knowledge biased segmentation for osteoporosis detection from X-ray, dual X-ray and CT images. *Curr Med Imaging*. 2021;17(2):288-295. [doi: [10.2174/1573405616999200730175526](https://doi.org/10.2174/1573405616999200730175526)] [Medline: [32748751](#)]
20. Yu X, Ye C, Xiang L. Application of artificial neural network in the diagnostic system of osteoporosis. *Neurocomputing*. Nov 2016;214:376-381. [doi: [10.1016/j.neucom.2016.06.023](https://doi.org/10.1016/j.neucom.2016.06.023)]
21. Dzierżak R, Omiotek Z. Application of deep convolutional neural networks in the diagnosis of osteoporosis. *Sensors (Basel)*. Oct 26, 2022;22(21):8189. [doi: [10.3390/s22218189](https://doi.org/10.3390/s22218189)] [Medline: [36365886](#)]
22. Chen B, Cui J, Li C, et al. Application of radiomics model based on lumbar computed tomography in diagnosis of elderly osteoporosis. *J Orthop Res*. Jun 2024;42(6):1356-1368. [doi: [10.1002/jor.25789](https://doi.org/10.1002/jor.25789)] [Medline: [38245854](#)]
23. Tong X, Wang S, Zhang J, Fan Y, Liu Y, Wei W. Automatic osteoporosis screening system using radiomics and deep learning from low-dose chest CT images. *Bioengineering (Basel)*. Jan 2, 2024;11(1):50. [doi: [10.3390/bioengineering11010050](https://doi.org/10.3390/bioengineering11010050)] [Medline: [38247927](#)]
24. Dhanagopal R, Menaka R, Suresh Kumar R, Vasanth Raj PT, Debrah EL, Pradeep K. Channel-boosted and transfer learning convolutional neural network-based osteoporosis detection from CT scan, dual X-ray, and X-ray images. *J Healthc Eng*. 2024;2024(3733705):3733705. [doi: [10.1155/2024/3733705](https://doi.org/10.1155/2024/3733705)] [Medline: [38223259](#)]
25. Breit HC, Varga-Szemes A, Schoepf UJ, et al. CNN-based evaluation of bone density improves diagnostic performance to detect osteopenia and osteoporosis in patients with non-contrast chest CT examinations. *Eur J Radiol*. Apr 2023;161:110728. [doi: [10.1016/j.ejrad.2023.110728](https://doi.org/10.1016/j.ejrad.2023.110728)] [Medline: [36773426](#)]
26. Tang C, Zhang W, Li H, et al. CNN-based qualitative detection of bone mineral density via diagnostic CT slices for osteoporosis screening. *Osteoporos Int*. May 2021;32(5):971-979. [doi: [10.1007/s00198-020-05673-w](https://doi.org/10.1007/s00198-020-05673-w)]
27. Mao L, Xia Z, Pan L, et al. Deep learning for screening primary osteopenia and osteoporosis using spine radiographs and patient clinical covariates in a Chinese population. *Front Endocrinol (Lausanne)*. 2022;13(971877):971877. [doi: [10.3389/fendo.2022.971877](https://doi.org/10.3389/fendo.2022.971877)] [Medline: [36176468](#)]
28. Kim S, Kim BR, Chae HD, et al. Deep radiomics-based approach to the diagnosis of osteoporosis using hip radiographs. *Radiol Artif Intell*. Jul 2022;4(4):e210212. [doi: [10.1148/ryai.210212](https://doi.org/10.1148/ryai.210212)] [Medline: [35923378](#)]
29. Zhang K, Lin PC, Pan J, et al. DeepmdQCT: a multitask network with domain invariant features and comprehensive attention mechanism for quantitative computer tomography diagnosis of osteoporosis. *Comput Biol Med*. Mar 2024;170:107916. [doi: [10.1016/j.combiomed.2023.107916](https://doi.org/10.1016/j.combiomed.2023.107916)] [Medline: [38237237](#)]
30. Zhang B, Chen Z, Yan R, et al. Development and validation of a feature-based broad-learning system for opportunistic osteoporosis screening using lumbar spine radiographs. *Acad Radiol*. Jan 2024;31(1):84-92. [doi: [10.1016/j.acra.2023.07.002](https://doi.org/10.1016/j.acra.2023.07.002)]
31. Niu X, Huang Y, Li X, et al. Development and validation of a fully automated system using deep learning for opportunistic osteoporosis screening using low-dose computed tomography scans. *Quant Imaging Med Surg*. Aug 1, 2023;13(8):5294-5305. [doi: [10.21037/qims-22-1438](https://doi.org/10.21037/qims-22-1438)] [Medline: [37581046](#)]
32. Xie Q, Chen Y, Hu Y, et al. Development and validation of a machine learning-derived radiomics model for diagnosis of osteoporosis and osteopenia using quantitative computed tomography. *BMC Med Imaging*. Aug 8, 2022;22(1):140. [doi: [10.1186/s12880-022-00868-5](https://doi.org/10.1186/s12880-022-00868-5)] [Medline: [35941568](#)]
33. Uemura K, Otake Y, Takashima K, et al. Development and validation of an open-source tool for opportunistic screening of osteoporosis from hip CT images. *Bone Joint Res*. Sep 20, 2023;12(9):590-597. [doi: [10.1302/2046-3758.129.BJR-2023-0115.R1](https://doi.org/10.1302/2046-3758.129.BJR-2023-0115.R1)] [Medline: [37728034](#)]
34. Zaman MU, Alam MK, Alqhtani NR, et al. RETRACTED ARTICLE: Diagnosing osteoporosis using deep neural network-assisted optical image processing method. *Opt Quant Electron*. Mar 2024;56(3). [doi: [10.1007/s11082-023-06031-w](https://doi.org/10.1007/s11082-023-06031-w)]

35. Lee JH, Hwang YN, Park SY, Jeong JH, Kim SM. Diagnosis of osteoporosis by quantification of trabecular microarchitectures from hip radiographs using artificial neural networks. *J Comp Theo Nano*. Jul 1, 2015;12(7):1115-1120. [doi: [10.1166/jctn.2015.3859](https://doi.org/10.1166/jctn.2015.3859)]

36. Kavitha MS, Asano A, Taguchi A, Kurita T, Sanada M. Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. *BMC Med Imaging*. Jan 16, 2012;12(1):22248480. [doi: [10.1186/1471-2342-12-1](https://doi.org/10.1186/1471-2342-12-1)] [Medline: [22248480](#)]

37. Yamamoto N, Sukegawa S, Yamashita K, et al. Effect of patient clinical variables in osteoporosis classification using hip X-rays in deep learning analysis. *Medicina (Kaunas)*. Aug 20, 2021;57(8):846. [doi: [10.3390/medicina57080846](https://doi.org/10.3390/medicina57080846)] [Medline: [34441052](#)]

38. Pan J, Lin PC, Gong SC, et al. Effectiveness of opportunistic osteoporosis screening on chest CT using the DCNN model. *BMC Musculoskelet Disord*. Feb 27, 2024;25(1):176. [doi: [10.1186/s12891-024-07297-1](https://doi.org/10.1186/s12891-024-07297-1)] [Medline: [38413868](#)]

39. Zhang K, Lin P, Pan J, et al. End to end multitask joint learning model for osteoporosis classification in CT images. *Comput Intell Neurosci*. 2023;2023:3018320. [doi: [10.1155/2023/3018320](https://doi.org/10.1155/2023/3018320)] [Medline: [36970245](#)]

40. Oh J, Kim B, Oh G, Hwangbo Y, Ye JC. End-to-end semi-supervised opportunistic osteoporosis screening using computed tomography. *Endocrinol Metab (Seoul)*. Jun 2024;39(3):500-510. [doi: [10.3803/EnM.2023.1860](https://doi.org/10.3803/EnM.2023.1860)] [Medline: [38721637](#)]

41. Oh S, Kang WY, Park H, et al. Evaluation of deep learning-based quantitative computed tomography for opportunistic osteoporosis screening. *Sci Rep*. Jan 5, 2024;14(1):363. [doi: [10.1038/s41598-023-45824-7](https://doi.org/10.1038/s41598-023-45824-7)] [Medline: [38182616](#)]

42. Lee KS, Jung SK, Ryu JJ, Shin SW, Choi J. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med*. Feb 1, 2020;9(2):392. [doi: [10.3390/jcm9020392](https://doi.org/10.3390/jcm9020392)] [Medline: [32024114](#)]

43. Xu Y, Li D, Chen Q, Fan Y. Full supervised learning for osteoporosis diagnosis using micro-CT images. *Microsc Res Tech*. Apr 2013;76(4):333-341. [doi: [10.1002/jemt.22171](https://doi.org/10.1002/jemt.22171)]

44. Wang S, Tong X, Cheng Q, et al. Fully automated deep learning system for osteoporosis screening using chest computed tomography images. *Quant Imaging Med Surg*. Apr 3, 2024;14(4):2816-2827. [doi: [10.21037/qims-23-1617](https://doi.org/10.21037/qims-23-1617)] [Medline: [38617137](#)]

45. Zhao Y, Zhao T, Chen S, et al. Fully automated radiomic screening pipeline for osteoporosis and abnormal bone density with a deep learning-based segmentation using a short lumbar mDixon sequence. *Quant Imaging Med Surg*. Feb 2022;12(2):1198-1213. [doi: [10.21037/qims-21-587](https://doi.org/10.21037/qims-21-587)] [Medline: [35111616](#)]

46. Su R, Liu T, Sun C, Jin Q, Jennane R, Wei L. Fusing convolutional neural network features with hand-crafted features for osteoporosis diagnoses. *Neurocomputing*. Apr 2020;385:300-309. [doi: [10.1016/j.neucom.2019.12.083](https://doi.org/10.1016/j.neucom.2019.12.083)]

47. Pickhardt PJ, Nguyen T, Perez AA, et al. Improved CT-based osteoporosis assessment with a fully automated deep learning tool. *Radiol Artif Intell*. Sep 2022;4(5):e220042. [doi: [10.1148/ryai.220042](https://doi.org/10.1148/ryai.220042)] [Medline: [36204542](#)]

48. Jiang C, Jin D, Ni M, Zhang Y, Yuan H. Influence of image reconstruction kernel on computed tomography-based finite element analysis in the clinical opportunistic screening of osteoporosis—a preliminary result. *Front Endocrinol*. 2023;14(1076990):36936156. [doi: [10.3389/fendo.2023.1076990](https://doi.org/10.3389/fendo.2023.1076990)]

49. Alshamrani K, Alshamrani HA. Lossless compression-based detection of osteoporosis using bone X-ray imaging. *J Xray Sci Technol*. 2024;32(2):475-491. [doi: [10.3233/XST-230238](https://doi.org/10.3233/XST-230238)] [Medline: [38393881](#)]

50. Sebro R, Elmahdy M. Machine learning for opportunistic screening for osteoporosis and osteopenia using knee CT scans. *Can Assoc Radiol J*. Nov 2023;74(4):676-687. [doi: [10.1177/08465371231164743](https://doi.org/10.1177/08465371231164743)] [Medline: [36960893](#)]

51. Sebro R, De la Garza-Ramos C. Machine learning for opportunistic screening for osteoporosis from CT scans of the wrist and forearm. *Diagnostics (Basel)*. Mar 11, 2022;12(3):691. [doi: [10.3390/diagnostics12030691](https://doi.org/10.3390/diagnostics12030691)] [Medline: [35328244](#)]

52. Namatevs I, Nikulins A, Edelmers E, et al. Modular neural networks for osteoporosis detection in mandibular cone-beam computed tomography scans. *Tomography*. Sep 22, 2023;9(5):1772-1786. [doi: [10.3390/tomography9050141](https://doi.org/10.3390/tomography9050141)] [Medline: [37888733](#)]

53. Hwang DH, Bak SH, Ha TJ, Kim Y, Kim WJ, Choi HS. Multi-view computed tomography network for osteoporosis classification. *IEEE Access*. 2023;11:22297-22306. [doi: [10.1109/ACCESS.2023.3252361](https://doi.org/10.1109/ACCESS.2023.3252361)]

54. Fang Y, Li W, Chen X, et al. Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. *Eur Radiol*. Apr 2021;31(4):1831-1842. [doi: [10.1007/s00330-020-07312-8](https://doi.org/10.1007/s00330-020-07312-8)]

55. Yang J, Liao M, Wang Y, et al. Opportunistic osteoporosis screening using chest CT with artificial intelligence. *Osteoporos Int*. Dec 2022;33(12):2547-2561. [doi: [10.1007/s00198-022-06491-y](https://doi.org/10.1007/s00198-022-06491-y)]

56. Jang M, Kim M, Bae SJ, Lee SH, Koh JM, Kim N. Opportunistic osteoporosis screening using chest radiographs with deep learning: development and external validation with a cohort dataset. *J Bone Miner Res*. Feb 2022;37(2):369-377. [doi: [10.1002/jbmr.4477](https://doi.org/10.1002/jbmr.4477)] [Medline: [34812546](#)]

57. Sebro R, De la Garza-Ramos C. Opportunistic screening for osteoporosis and osteopenia from CT scans of the abdomen and pelvis using machine learning. *Eur Radiol*. Mar 2023;33(3):1812-1823. [doi: [10.1007/s00330-022-09136-0](https://doi.org/10.1007/s00330-022-09136-0)] [Medline: [36166085](#)]

58. Elmahdy M, Sebro R. Opportunistic screening for osteoporosis using CT scans of the knee: a pilot study. *Stud Health Technol Inform*. May 18, 2023;302:909-910. [doi: [10.3233/SHTI230305](https://doi.org/10.3233/SHTI230305)] [Medline: [37203533](#)]

59. Mohammadi FG, Sebro R. Opportunistic screening for osteoporosis using hand radiographs: a preliminary study. *Stud Health Technol Inform*. May 18, 2023;302:911-912. [doi: [10.3233/SHTI230306](https://doi.org/10.3233/SHTI230306)] [Medline: [37203534](#)]

60. Lee JS, Adhikari S, Liu L, Jeong HG, Kim H, Yoon SJ. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofac Radiol*. Jan 2019;48(1):20170344. [doi: [10.1259/dmfr.20170344](https://doi.org/10.1259/dmfr.20170344)] [Medline: [30004241](#)]

61. Wani IM, Arora S. Osteoporosis diagnosis in knee X-rays by transfer learning based on convolution neural network. *Multimed Tools Appl*. 2023;82(9):14193-14217. [doi: [10.1007/s11042-022-13911-y](https://doi.org/10.1007/s11042-022-13911-y)] [Medline: [36185321](#)]

62. Nakamoto T, Taguchi A, Kakimoto N. Osteoporosis screening support system from panoramic radiographs using deep learning by convolutional neural network. *Dentomaxillofac Radiol*. Sep 1, 2022;51(6):20220135. [doi: [10.1259/dmfr.20220135](https://doi.org/10.1259/dmfr.20220135)] [Medline: [35816516](#)]

63. Muramatsu C, Horiba K, Hayashi T, et al. Quantitative assessment of mandibular cortical erosion on dental panoramic radiographs for screening osteoporosis. *Int J Comput Assist Radiol Surg*. Nov 2016;11(11):2021-2032. [doi: [10.1007/s11548-016-1438-8](https://doi.org/10.1007/s11548-016-1438-8)] [Medline: [27289239](#)]

64. Kang S ru, Wang K. Radiomic nomogram based on lumbar spine magnetic resonance images to diagnose osteoporosis. *Acta Radiol*. Aug 2024;65(8):950-958. [doi: [10.1177/02841851241242052](https://doi.org/10.1177/02841851241242052)]

65. Jiang YW, Xu XJ, Wang R, Chen CM. Radiomics analysis based on lumbar spine CT to detect osteoporosis. *Eur Radiol*. Nov 2022;32(11):8019-8026. [doi: [10.1007/s00330-022-08805-4](https://doi.org/10.1007/s00330-022-08805-4)] [Medline: [35499565](#)]

66. He L, Liu Z, Liu C, et al. Radiomics based on lumbar spine magnetic resonance imaging to detect osteoporosis. *Acad Radiol*. Jun 2021;28(6):e165-e171. [doi: [10.1016/j.acra.2020.03.046](https://doi.org/10.1016/j.acra.2020.03.046)] [Medline: [32386949](#)]

67. Zhang H, Wei W, Qian B, et al. Screening for osteoporosis based on IQon spectral CT virtual low monoenergetic images: comparison with conventional 120 kVp images. *Heliyon*. Oct 2023;9(10):e20750. [doi: [10.1016/j.heliyon.2023.e20750](https://doi.org/10.1016/j.heliyon.2023.e20750)]

68. Krishnaraj A, Barrett S, Bregman-Amitai O, et al. Simulating dual-energy X-ray absorptiometry in CT using deep-learning segmentation cascade. *J Am Coll Radiol*. Oct 2019;16(10):1473-1479. [doi: [10.1016/j.jacr.2019.02.033](https://doi.org/10.1016/j.jacr.2019.02.033)]

69. Sebro R, De la Garza-Ramos C. Support vector machines are superior to principal components analysis for selecting the optimal bones' CT attenuations for opportunistic screening for osteoporosis using CT scans of the foot or ankle. *Osteoporos Sarcopenia*. Sep 2022;8(3):112-122. [doi: [10.1016/j.afos.2022.09.002](https://doi.org/10.1016/j.afos.2022.09.002)] [Medline: [36268496](#)]

70. Kavitha MS, An SY, An CH, et al. Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in Korean women. *Oral Surg Oral Med Oral Pathol Oral Radiol*. Mar 2015;119(3):346-356. [doi: [10.1016/j.oooo.2014.11.009](https://doi.org/10.1016/j.oooo.2014.11.009)] [Medline: [25600978](#)]

71. Kavitha MS, Asano A, Taguchi A, Heo MS. The combination of a histogram-based clustering algorithm and support vector machine for the diagnosis of osteoporosis. *Imaging Sci Dent*. 2013;43(3):153. [doi: [10.5624/isd.2013.43.3.153](https://doi.org/10.5624/isd.2013.43.3.153)]

72. Fang K, Zheng X, Lin X, Dai Z. Unveiling osteoporosis through radiomics analysis of hip CT imaging. *Acad Radiol*. Mar 2024;31(3):1003-1013. [doi: [10.1016/j.acra.2023.10.009](https://doi.org/10.1016/j.acra.2023.10.009)]

73. Xue Z, Huo J, Sun X, et al. Using radiomic features of lumbar spine CT images to differentiate osteoporosis from normal bone density. *BMC Musculoskelet Disord*. Dec 2022;23(1):35395769. [doi: [10.1186/s12891-022-05309-6](https://doi.org/10.1186/s12891-022-05309-6)]

74. Chen M, Gerges M, Raynor WY, et al. State of the art imaging of osteoporosis. *Semin Nucl Med*. May 2024;54(3):415-426. [doi: [10.1053/j.semnuclmed.2023.10.008](https://doi.org/10.1053/j.semnuclmed.2023.10.008)]

75. El Maghraoui A, Roux C. DXA scanning in clinical practice. *QJM*. Aug 2008;101(8):605-617. [doi: [10.1093/qjmed/hcn022](https://doi.org/10.1093/qjmed/hcn022)] [Medline: [18334497](#)]

76. Kessenich CR. Diagnostic imaging and biochemical markers of bone turnover. *Nurs Clin North Am*. Sep 2001;36(3):409-416, [Medline: [11532656](#)]

77. Sollmann N, Löffler MT, Kronthaler S, et al. MRI-based quantitative osteoporosis imaging at the spine and femur. *Magn Reson Imaging*. Jul 2021;54(1):12-35. [doi: [10.1002/jmri.27260](https://doi.org/10.1002/jmri.27260)]

78. Oei L, Koromani F, Rivadeneira F, Zillikens MC, Oei EHG. Quantitative imaging methods in osteoporosis. *Quant Imaging Med Surg*. Dec 2016;6(6):680-698. [doi: [10.21037/qims.2016.12.13](https://doi.org/10.21037/qims.2016.12.13)] [Medline: [28090446](#)]

79. Yang Q, Cheng H, Qin J, et al. A machine learning-based Preclinical Osteoporosis Screening Tool (POST): model development and validation study. *JMIR Aging*. Nov 8, 2023;6:e46791. [doi: [10.2196/46791](https://doi.org/10.2196/46791)] [Medline: [37986117](#)]

80. Baik SM, Kwon HJ, Kim Y, Lee J, Park YH, Park DJ. Machine learning model for osteoporosis diagnosis based on bone turnover markers. *Health Informatics J*. 2024;30(3):39115269. [doi: [10.1177/14604582241270778](https://doi.org/10.1177/14604582241270778)] [Medline: [39115269](#)]

81. Ou Yang WY, Lai CC, Tsou MT, Hwang LC. Development of machine learning models for prediction of osteoporosis from clinical health examination data. *Int J Environ Res Public Health*. Jul 18, 2021;18(14):7635. [doi: [10.3390/ijerph18147635](https://doi.org/10.3390/ijerph18147635)] [Medline: [34300086](https://pubmed.ncbi.nlm.nih.gov/34300086/)]
82. Putra RH, Doi C, Yoda N, Astuti ER, Sasaki K. Current applications and development of artificial intelligence for digital dental radiography. *Dentomaxillofac Radiol*. Jan 1, 2022;51(1):20210197. [doi: [10.1259/dmfr.20210197](https://doi.org/10.1259/dmfr.20210197)] [Medline: [34233515](https://pubmed.ncbi.nlm.nih.gov/34233515/)]
83. Du X, Chen Y, Zhao J, Xi Y. A convolutional neural network based auto-positioning method for dental arch in rotational panoramic radiography. *Annu Int Conf IEEE Eng Med Biol Soc*. Jul 2018;2018(2615-8):2615-2618. [doi: [10.1109/EMBC.2018.8512732](https://doi.org/10.1109/EMBC.2018.8512732)] [Medline: [30440944](https://pubmed.ncbi.nlm.nih.gov/30440944/)]
84. Liang K, Zhang L, Yang H, Yang Y, Chen Z, Xing Y. Metal artifact reduction for practical dental computed tomography by improving interpolation-based reconstruction with deep learning. *Med Phys*. Dec 2019;46(12):e823-e834. [doi: [10.1002/mp.13644](https://doi.org/10.1002/mp.13644)] [Medline: [31811792](https://pubmed.ncbi.nlm.nih.gov/31811792/)]

## Abbreviations

**AI:** artificial intelligence  
**CAD:** computer-aided diagnosis  
**CT:** computed tomography  
**DL:** deep learning  
**DOR:** diagnostic odds ratio  
**DXA:** dual-energy x-ray absorptiometry  
**ML:** machine learning  
**MRI:** magnetic resonance imaging  
**NLR:** negative likelihood ratio  
**OP:** osteoporosis  
**PLR:** positive likelihood ratio  
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
**SEN:** sensitivity  
**SPC:** specificity  
**SROC:** summary receiver operating characteristic

Edited by Javad Sarvestan; peer-reviewed by Mincong He, Natthapong Nanthasamroeng; submitted 14.Apr.2025; final revised version received 03.Jun.2025; accepted 04.Jun.2025; published 16.Jan.2026

Please cite as:

Zhao R, Yang H, Li Y, Li X, Yang Z, Lin Y, Huang J, Wan L, Huang H  
*The Diagnostic Value of Image-Based Machine Learning for Osteoporosis: Systematic Review and Meta-Analysis*  
*J Med Internet Res* 2026;28:e75965  
URL: <https://www.jmir.org/2026/1/e75965>  
doi: [10.2196/75965](https://doi.org/10.2196/75965)

© Rui Zhao, Haolin Yang, Yangbo Li, Xiaoyun Li, Zhijie Yang, Yanping Lin, Jiachun Huang, Lei Wan, Hongxing Huang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 16.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.