



Fully automated deep learning system for osteoporosis screening using chest computed tomography images

Shigeng Wang^{1#^}, Xiaoyu Tong^{1#^}, Qiye Cheng¹, Qingzhu Xiao², Jingjing Cui³, Jianying Li⁴, Yijun Liu¹, Xin Fang^{1^}

¹Department of Radiology, The First Affiliated Hospital of Dalian Medical University, Dalian, China; ²School of Investment and Project Management, Dongbei University of Finance and Economics, Dalian, China; ³United Imaging Intelligence, Beijing, China; ⁴CT Research, GE Healthcare, Dalian, China

Contributions: (I) Conception and design: S Wang, X Tong; (II) Administrative support: Y Liu, X Fang; (III) Provision of study materials or patients: Q Cheng, X Tong; (IV) Collection and assembly of data: S Wang, X Fang; (V) Data analysis and interpretation: S Wang, X Tong, Q Xiao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

Correspondence to: Xin Fang, MS. Department of Radiology, The First Affiliated Hospital of Dalian Medical University, Xigang district, 193 Lianhe Road, Dalian 116011, China. Email: FX25632331@163.com.

Background: Osteoporosis, a disease stemming from bone metabolism irregularities, affects approximately 200 million people worldwide. Timely detection of osteoporosis is pivotal in grappling with this public health challenge. Deep learning (DL), emerging as a promising methodology in the field of medical imaging, holds considerable potential for the assessment of bone mineral density (BMD). This study aimed to propose an automated DL framework for BMD assessment that integrates localization, segmentation, and ternary classification using various dominant convolutional neural networks (CNNs).

Methods: In this retrospective study, a cohort of 2,274 patients underwent chest computed tomography (CT) was enrolled from January 2022 to June 2023 for the development of the integrated DL system. The study unfolded in 2 phases. Initially, 1,025 patients were selected based on specific criteria to develop an automated segmentation model, utilizing 2 VB-Net networks. Subsequently, a distinct cohort of 902 patients was employed for the development and testing of classification models for BMD assessment. Then, 3 distinct DL network architectures, specifically DenseNet, ResNet-18, and ResNet-50, were applied to formulate the 3-classification BMD assessment model. The performance of both phases was evaluated using an independent test set consisting of 347 individuals. Segmentation performance was evaluated using the Dice similarity coefficient; classification performance was appraised using the receiver operating characteristic (ROC) curve. Furthermore, metrics such as the area under the curve (AUC), accuracy, and precision were meticulously calculated.

Results: In the first stage, the automatic segmentation model demonstrated excellent segmentation performance, with mean Dice surpassing 0.93 in the independent test set. In the second stage, both the DenseNet and ResNet-18 demonstrated excellent diagnostic performance in detecting bone status. For osteoporosis, and osteopenia, the AUCs were as follows: DenseNet achieved 0.94 [95% confidence interval (CI): 0.91–0.97], and 0.91 (95% CI: 0.87–0.94), respectively; ResNet-18 attained 0.96 (95% CI: 0.92–0.98), and 0.91 (95% CI: 0.87–0.94), respectively. However, the ResNet-50 model exhibited suboptimal diagnostic performance for osteopenia, with an AUC value of only 0.76 (95% CI: 0.69–0.80). Alterations in tube voltage had a more pronounced impact on the performance of the DenseNet. In the independent test set with tube

[^] ORCID: Shigeng Wang, 0000-0002-5643-9026; Xiaoyu Tong, 0000-0003-3814-124X; Xin Fang, 0000-0002-1125-3658.

voltage at 100 kVp images, the accuracy and precision of DenseNet decreased on average by approximately 14.29% and 18.82%, respectively, whereas the accuracy and precision of ResNet-18 decreased by about 8.33% and 7.14%, respectively.

Conclusions: The state-of-the-art DL framework model offers an effective and efficient approach for opportunistic osteoporosis screening using chest CT, without incurring additional costs or radiation exposure.

Keywords: Bone mineral density (BMD); osteoporosis; deep learning (DL); computed tomography (CT)

Submitted Nov 14, 2023. Accepted for publication Feb 21, 2024. Published online Mar 21, 2024.

doi: 10.21037/qims-23-1617

View this article at: <https://dx.doi.org/10.21037/qims-23-1617>

Introduction

Osteoporosis, a disease of bone metabolism, affects approximately 200 million individuals globally (1). The pathology is characterized by a reduction in bone mineral density (BMD) and degeneration of bone trabeculae, leading to an increased risk of fractures (2). In China, the incidence of osteoporotic fractures is projected to rise from 3.5 million in 2010 to 4.5 million in 2025 with the aging population, representing a 28% surge (3). These fractures not only contribute to morbidity, dysfunction, and diminished quality of life but also impose a substantial burden on the healthcare system (4). Early detection of osteoporosis is pivotal in addressing this public health challenge. Currently, BMD is measured using dual-energy X-ray absorptiometry (DXA) and quantitative computed tomography (QCT) in clinical practice (5-7). However, both methods are underutilized due to intricate post-processing techniques, high equipment costs, and a shortage of skilled operators. Only 19–37% of eligible Medicare beneficiaries in the United States undergo BMD testing (8,9).

In contrast, the utilization of CT scans, particularly for chest imaging, is consistently high and on the rise. Chest CT scans are frequently recommended for detecting emphysema and screening for lung cancer, with more than 20 million chest CT examinations performed annually in the United States alone (10,11). The comprehensive integration and utilization of BMD information from extensive chest CT data is an appealing prospect, offering the potential to facilitate opportunistic osteoporosis screening without incurring additional exposure and high costs.

In recent years, deep learning (DL) has emerged as a highly effective machine learning technique for improving computerized image recognition through the utilization of multilayer neural networks (12). Consequently, numerous

research groups have proposed methods for opportunistic osteoporosis screening by leveraging pre-existing images, achieving commendable performance (13-15). However, these approaches have predominantly concentrated on osteoporosis detection, overlooking the crucial aspects of vertebral body location and segmentation, thereby imposing a significant burden on radiologists. Furthermore, existing methods treat osteoporosis as a binary problem, neglecting the urgent need and strong incentive to transform it into a trinomial problem encompassing osteoporosis, osteopenia, and normal BMD (16). Although the classification of these 3 categories is more challenging, the inclusion of osteopenia can enhance predictability in the prevention and treatment of osteoporosis (15). To address these issues, we propose a comprehensive DL framework for BMD assessment that integrates localization, segmentation, and ternary classification using various state-of-the-art convolutional neural networks (CNNs). We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1617/rc>).

Methods

Participants

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Prior to the commencement of this retrospective study, ethical approval was secured from the Ethics Committee of the First Affiliated Hospital of Dalian Medical University (No. PJ-KS-KY-2023-276), and the requirement for individual consent for this retrospective analysis was waived. Patients who underwent chest CT scans between January 2022 and March 2023 were retrieved from the picture archiving and

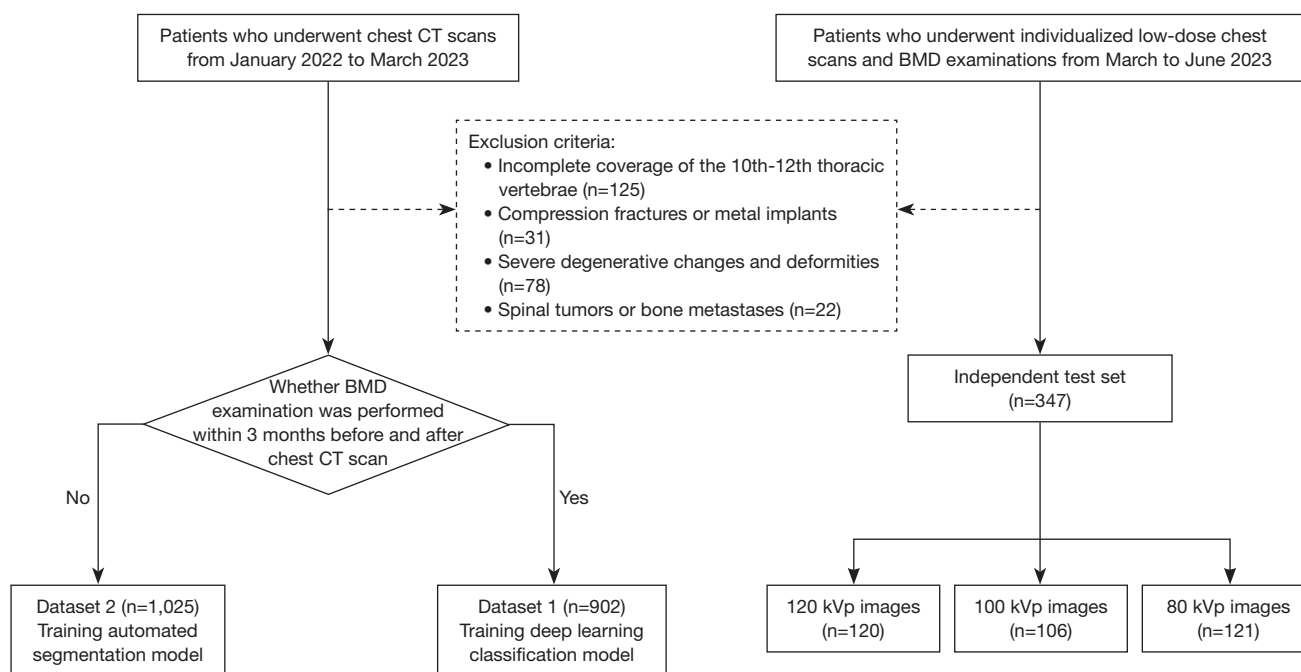


Figure 1 Flowchart of patient recruitment. CT, computed tomography; BMD, bone mineral density.

communication system. Exclusions were made for cases with incomplete coverage of the 10th–12th thoracic vertebrae in the chest CT scans, as well as for those with compression fractures, metal implants, severe degenerative changes, deformities, spinal tumors, or bone metastases. Bone status remains relatively stable within 3 months, so patients who underwent a BMD examination within 3 months before or after the chest CT scan were allocated to Dataset 1, and further divided into a training set and an internal test set at an 8:2 ratio for constructing and validating a DL system for comprehensive BMD assessment. The remaining patients were incorporated into Dataset 2 for training an automatic segmentation model. Additionally, patients who underwent individualized low-dose chest scans and BMD examinations (physical examination item) from March to June 2023 were included, labeled as an independent test to evaluate the performance of the DL system in segmentation and classification. *Figure 1* provides a detailed illustration of the inclusion process.

CT image acquisition

All participants underwent a chest CT scan using a 256-row detector CT system (Revolution CT; GE, Waukesha, WI, USA) covering the range from the apical lung to 2 cm

below the diaphragm. The standard chest CT scan followed a standard scanning protocol with a fixed tube voltage of 120 kVp and Smart mA. For the individualized low-dose chest scan, automatic tube voltage selection technology was utilized, whereas the remaining scanning parameters were maintained consistently: a tube rotation time of 0.5 s/r, a pitch of 0.992:1, a matrix size of 512×512, and a slice thickness and interval of 5 mm. Image reconstruction was performed with a slice thickness and interval of 1.25 mm, employing a standard convolution kernel (std).

BMD examination and measurement

All BMD examinations adhered to the standardized protocol recommended by Mindways software (Mindways Software, Austin, TX, USA). Calibration phantoms underwent asynchronous scanning once a week to ensure quality control and measurement accuracy. Detailed information on the BMD measurement method can be found in the supplementary material ([Appendix 1](#)). In accordance with the BMD diagnostic guidelines set forth by the American College of Radiology (17), BMD values below 80 mg/cm³ indicate osteoporosis, values ranging from 80 to 120 mg/cm³ indicate osteopenia, and BMD values exceeding 120 mg/cm³ indicate normal BMD.

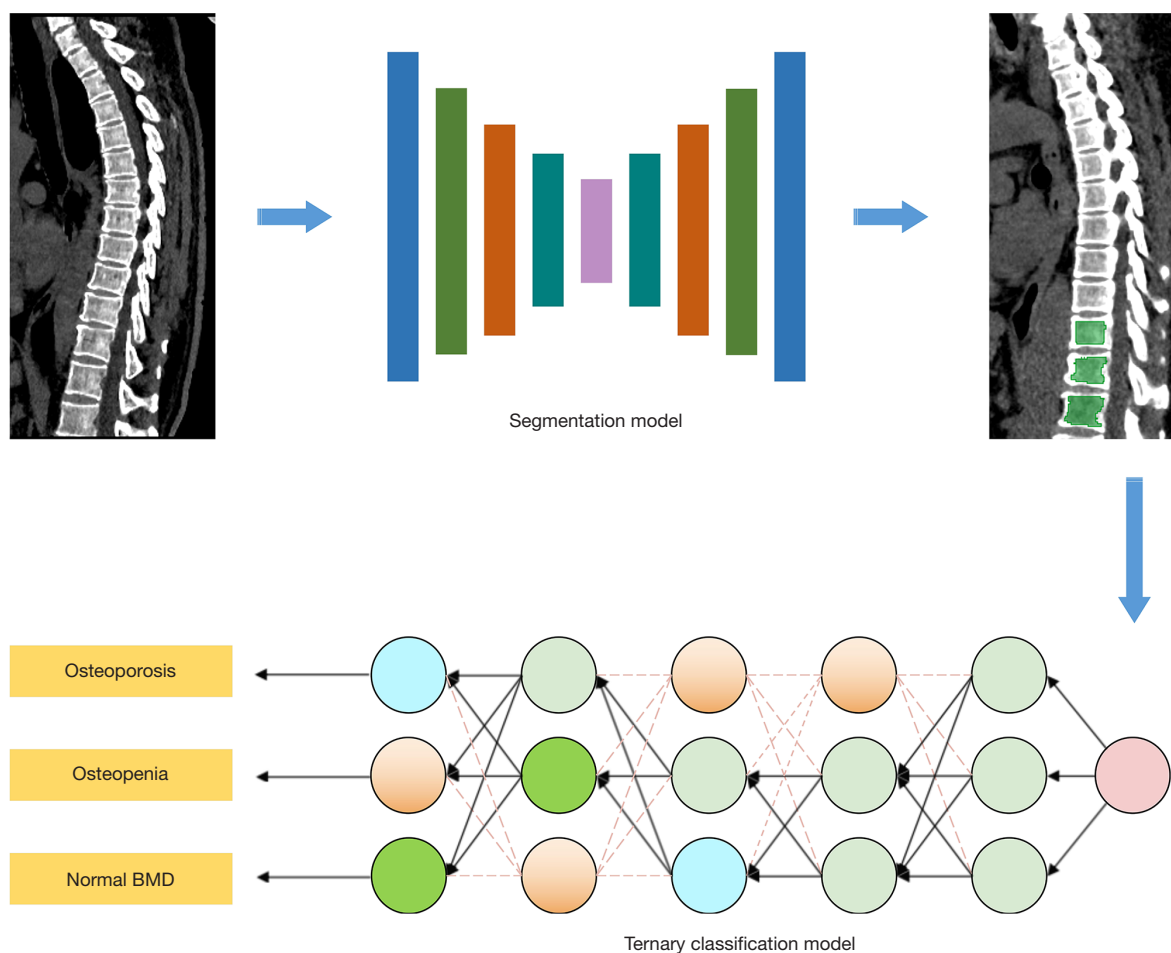


Figure 2 The overall pipeline of this study. BMD, bone mineral density.

Development of a fully automated DL system

The development of this fully automated DL network for osteoporosis diagnosis comprised 2 primary components. Firstly, VB-Net network underwent training to automatically segment trabecular vertebral cancellous bone (TVCB) in the T10–T12 vertebrae. Subsequently, a ternary model for BMD assessment was trained utilizing DenseNet, ResNet-18, and ResNet-50 network architectures. All procedures in this study were executed on the uAI Research Portal V1.1 (Shanghai United Imaging Intelligence, Co., Ltd., Shanghai, China). *Figure 2* illustrates the workflow of the entire study.

Construction of an automatic segmentation model

The TVCB in the T10–T12 vertebrae of Dataset 2 was manually delineated and labeled by 2 experienced

radiologists, M. Hu and W. Wei, with 3 and 6 years of expertise in musculoskeletal radiology, respectively. The identification of target vertebrae occurred on the mid-sagittal plane, and a layer-by-layer delineation of the region of interest within TVCB was conducted on the transverse images. Special care was taken to exclude areas of abnormal density, such as cortical bone, bone islands, and vertebral venous plexus during the delineation process. To assess inter-observer consistency, 100 cases were randomly selected and simultaneously delineated by both observers.

The VB-Net network architecture was employed to train the fully automated segmentation model. It adopts a bottleneck structure, instead of the convolutional layers used in the traditional V-Net, reducing the model size and inference time. Prior studies have showcased the superior segmentation accuracy of VB-Net across various tissues, including thoracic organs, the brain, and the spine

(18–20). A multi-scale strategy was employed for model training. Initially, images were resampled to a resolution of 3×3×3 mm to train a VB-Net coarse-scale network for the localization and coarse segmentation of TVCB. Subsequently, images were further resampled to a resolution of 1×1×1 mm isotropic voxel to enable the VB-Net fine-scale network to precisely recognize the TVCB boundary. Finally, a cascading approach was utilized to combine the outputs of the 2 VB-Nets, adhering to the principle of coarse-to-fine segmentation. The entire training process utilized a learning rate of $1e^{-4}$, a batch size of 8, 1,001 epochs, Adam optimizer, and Focal loss as the loss function. The performance of the automated segmentation model was evaluated using the Dice similarity coefficient between manual and automatic segmentation.

Construction of a ternary BMD assessment model

In Dataset 1, a ternary BMD assessment model was developed utilizing 3 distinct DL network architectures: DenseNet, ResNet-18, and ResNet-50. DenseNet employs a dense connectivity mechanism, where each layer receives input from all preceding layers, facilitating efficient gradient backpropagation and enhancing training efficiency. ResNet-18 and ResNet-50 are part of the ResNet learning network, incorporating skip connections to weakly link layers at intervals and mitigate strong dependencies between layers. The primary distinction between the 2 lies in their network depth and parameters. ResNet-18 comprises 2 3'3 convolutional networks connected as a ResNet block, whereas ResNet-50 is composed of bottleneck structures (1'1, 3'3, and 1'1) connected in sequence. During the training of the ternary DL network, all images undergo preprocessing steps involving 1 mm isotropic resampling and grayscale normalization. The learning rate is set to $1e^{-4}$, the batch size is fixed at 8, and the number of epochs is set to 101. To monitor model convergence, the Focal loss function and Adam optimizer are employed.

Statistical analysis

Statistical analyses were conducted using SPSS 24.0 (IBM Corp., Armonk, NY, USA) and MedCalc version 20.022 (MedCalc Ltd., Ostend, Belgium). Differences in categorical data among various groups were assessed using the chi-square test. One-way analysis of variance or the Kruskal-Wallis H test was employed to analyze the differences in continuous variables among groups, according

to normality. The diagnostic performance of the model was evaluated by constructing a receiver operating characteristic (ROC) curve, calculating metrics such as the area under the curve (AUC), F1 score, recall, precision, and accuracy. A significance level of $P < 0.05$ was considered statistically significant.

Results

Participant characteristics

The study included a total of 2,274 patients. Dataset 2, used for training the automatic segmentation model, comprised 1,025 participants [median age 64 years, interquartile range (IQR): 56–70 years; 576 males and 449 females]. Dataset 1, utilized for both training and testing the 3 DL classification BMD assessment models, included 902 patients [median age 64 years (IQR, 56–70 years); 480 males and 422 females]. The independent test set, aimed at evaluating the performance of the DL system, involved 347 patients [median age 65 years (IQR, 57–70 years); 209 males and 138 females]. Among them, 120 used a tube voltage of 120 kVp, 106 used a tube voltage of 100 kVp, and 121 used a tube voltage of 80 kVp. No statistically significant differences were observed in gender distribution, bone status, and age among the various cohorts. *Table 1* provides a summary of the BMD distribution as well as demographic characteristics for Dataset 1 and the independent test set.

Overall performance of the automatic DL system

Segmentation

The inter-observer consistency in manual segmentation exhibited a superior mean Dice coefficient of 0.92. The automated segmentation model demonstrated exceptional accuracy in identifying and segmenting TVCB, as evidenced by its outstanding performance in the independent test set [median Dice values: 0.95 (IQR, 0.93–0.97), with approximately 94% of participants achieving Dice values above 0.9]. The mean Dice coefficients for the segmentation model in 80, 100, and 120 kVp images of the independent test set were 0.93, 0.95, and 0.96, respectively. *Figure 3* illustrates the detailed Dice distribution.

Classification

The DL models based on DenseNet and ResNet-18 exhibited remarkable diagnostic performance in the internal test set. For osteoporosis, osteopenia, and normal

Table 1 The bone mineral density distribution as well as demographic characteristics for Dataset 1 and the independent test set

Characteristic	Training set (n=635)	Internal test set (n=267)	Independent test set (n=347)	P value
Gender, n (%)				0.063
Male	343 (54.02)	137 (51.31)	209 (60.23)	
Female	292 (45.98)	130 (48.69)	138 (39.77)	
Age, years, median [IQR]	64 [56–70]	63 [56–70]	65 [57–70]	0.491
Bone status, n (%)				0.335
Osteoporosis	190 (29.92)	77 (28.84)	82 (23.63)	
Osteopenia	217 (34.17)	93 (34.83)	128 (36.89)	
Normal BMD	228 (35.91)	97 (36.33)	137 (39.48)	

The gender and bone status are expressed as number (frequency). BMD, bone mineral density; IQR, interquartile range.

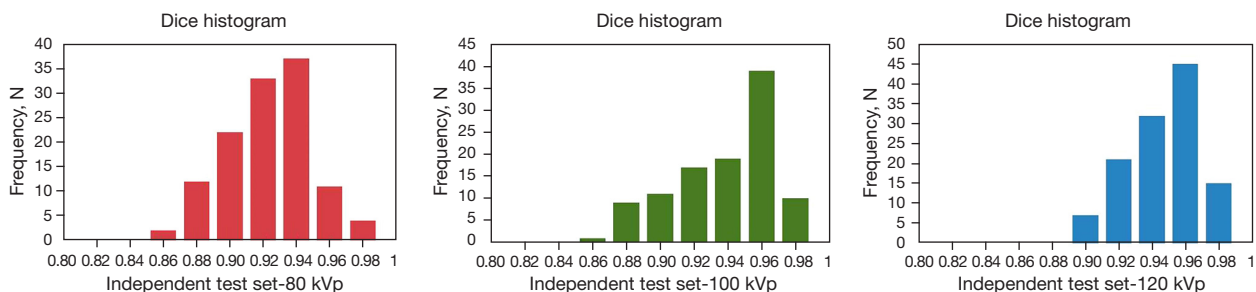


Figure 3 The Dice histograms of the independent test set with 80, 100, and 120 kVp. The mean Dice coefficients of three independent test sets of 80, 100, and 120 kVp were 0.93, 0.95, and 0.96, respectively, indicating excellent segmentation performance.

BMD, the AUCs were as follows: DenseNet achieved 0.94 [95% confidence interval (CI): 0.91–0.97], 0.91 (95% CI: 0.87–0.94), and 0.98 (95% CI: 0.96–0.99), respectively; ResNet-18 attained 0.96 (95% CI: 0.92–0.98), 0.91 (95% CI: 0.87–0.94), and 0.97 (95% CI: 0.94–0.99), respectively. The accuracy, F1 score, and recall rates for both models were 0.84 and 0.83, while the precision rates were 0.85 and 0.83. Conversely, the DL model based on ResNet-50 showed inferior diagnostic performance in BMD assessment, particularly in predicting osteopenia, with an AUC value of 0.76 (95% CI: 0.69–0.80). The accuracy, F1 score, recall rate and precision of this model were 0.67, 0.68, 0.69, and 0.71, respectively. Detailed results are presented in *Table 2* and *Figure 4*.

Gender-stratified analysis

In the internal test set, both the DL models based on DenseNet and ResNet-18 achieved AUC values exceeding 0.90 and accuracy surpassing 0.80 for males and females.

In contrast, the DL models based on ResNet-50 had AUC values of 0.90 for males and 0.87 for females, with corresponding accuracies of 0.73 and 0.61, respectively. Specific advantages of the diagnostic performance are outlined in *Table 3*.

Tube voltage analysis

Variations in tube voltage had a discernible impact on the BMD evaluation across all 3 DL models. Setting the tube voltage to 100 kVp resulted in an average decline in accuracy and precision of approximately 14.29% and 18.82%, respectively, for the DenseNet, and 8.33% and 7.14%, respectively, for ResNet-18. Employing 80 kVp images significantly degraded the performance of the DenseNet model, with an average decrease in accuracy and precision of 39.29% and 31.76%, respectively, compared to the use of 120 kVp images. Similarly, the ResNet-18 model exhibited an average decrease in accuracy and precision of 35.71% and 26.19%, respectively, compared

Table 2 The diagnostic performance of the constructed models for bone mineral density assessment

Sets	Images	Model	AUC	95% CI	Accuracy	F1 score	Recall	Precision
Internal test set	120 kVp	Model-DenseNet	0.95	0.93–0.97	0.84	0.84	0.84	0.85
		Model-ResNet-18	0.95	0.91–0.97	0.83	0.83	0.83	0.83
		Model-ResNet-50	0.89	0.85–0.92	0.67	0.68	0.69	0.71
Independent test set	120 kVp	Model-DenseNet	0.97	0.91–0.99	0.84	0.85	0.84	0.85
		Model-ResNet-18	0.97	0.91–0.99	0.84	0.85	0.86	0.84
		Model-ResNet-50	0.89	0.83–0.93	0.66	0.66	0.69	0.69
	100 kVp	Model-DenseNet	0.90	0.83–0.95	0.72	0.64	0.64	0.69
		Model-ResNet-18	0.93	0.86–0.98	0.77	0.76	0.75	0.78
		Model-ResNet-50	0.90	0.82–0.94	0.77	0.78	0.79	0.77
	80 kVp	Model-DenseNet	0.78	0.69–0.84	0.51	0.42	0.46	0.58
		Model-ResNet-18	0.84	0.76–0.89	0.54	0.47	0.5	0.62
		Model-ResNet-50	0.86	0.79–0.91	0.66	0.64	0.63	0.74

AUC, area under the curve; CI, confidence interval.

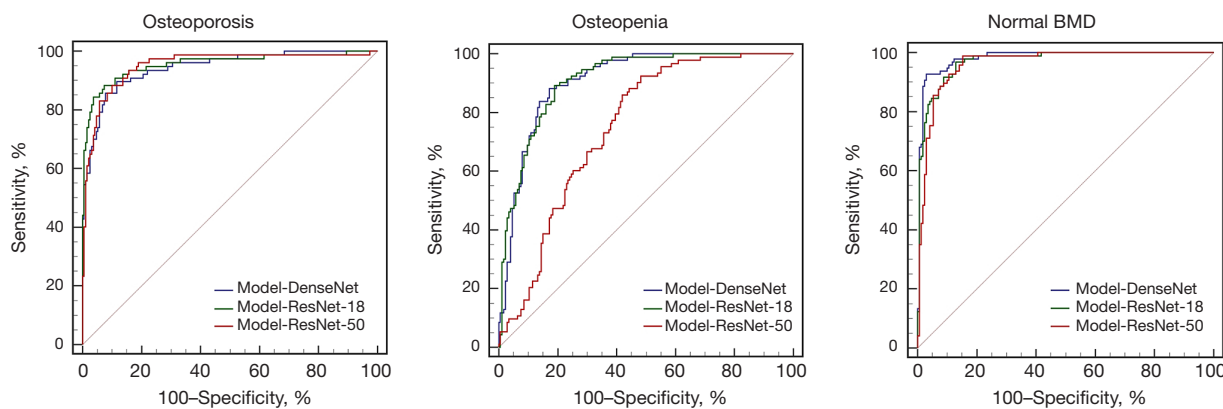


Figure 4 ROC analysis showing the diagnostic performance of these models for BMD assessment in the internal test set. The deep learning models using DenseNet (blue line), ResNet-18 (green line), and ResNet-50 (red line) achieved the following AUCs for osteoporosis, osteopenia, and normal BMD: DenseNet 0.94 (95% CI: 0.91–0.97), 0.91 (95% CI: 0.87–0.94), and 0.98 (95% CI: 0.96–0.99); ResNet-18 0.96 (95% CI: 0.92–0.98), 0.91 (95% CI: 0.87–0.94), and 0.97 (95% CI: 0.94–0.99); ResNet-50 0.95 (95% CI: 0.92–0.97), 0.76 (95% CI: 0.69–0.80), and 0.96 (95% CI: 0.93–0.98), respectively. DenseNet and ResNet-18 models exhibited superior performance compared to ResNet-50 model. ROC, receiver operating characteristic; BMD, bone mineral density; AUC, area under the curve; CI, confidence interval.

to using 120 kVp images. Conversely, the diagnostic performance of the ResNet-50 model was also influenced by the image tube voltage, but its overall performance was poorer, with accuracy values of 0.66, 0.77, and 0.66 for 80, 100, and 120 kVp images, respectively. Detailed diagnostic performance results for each model are presented in *Table 2* and *Figure 5*.

Discussion

This study endeavored to automate opportunistic osteoporosis screening through chest CT scanning, leveraging recent strides in DL within medical imaging (21). Initially, a rapid and precise automatic segmentation of TVCB was achieved using a VB-Net network, yielding a mean Dice coefficient of 0.95. This methodology

Table 3 The diagnostic performance of these models in gender-stratified analysis

Gender	Model	AUC	95% CI	Accuracy	F1 score	Recall	Precision
Male	Model-DenseNet	0.94	0.89–0.97	0.80	0.78	0.77	0.83
	Model-ResNet-18	0.95	0.90–0.98	0.85	0.86	0.85	0.87
	Model-ResNet-50	0.90	0.84–0.94	0.73	0.74	0.75	0.75
Female	Model-DenseNet	0.95	0.90–0.98	0.90	0.90	0.90	0.90
	Model-ResNet-18	0.94	0.89–0.97	0.80	0.80	0.80	0.80
	Model-ResNet-50	0.87	0.82–0.91	0.61	0.60	0.62	0.67

AUC, area under the curve; CI, confidence interval.

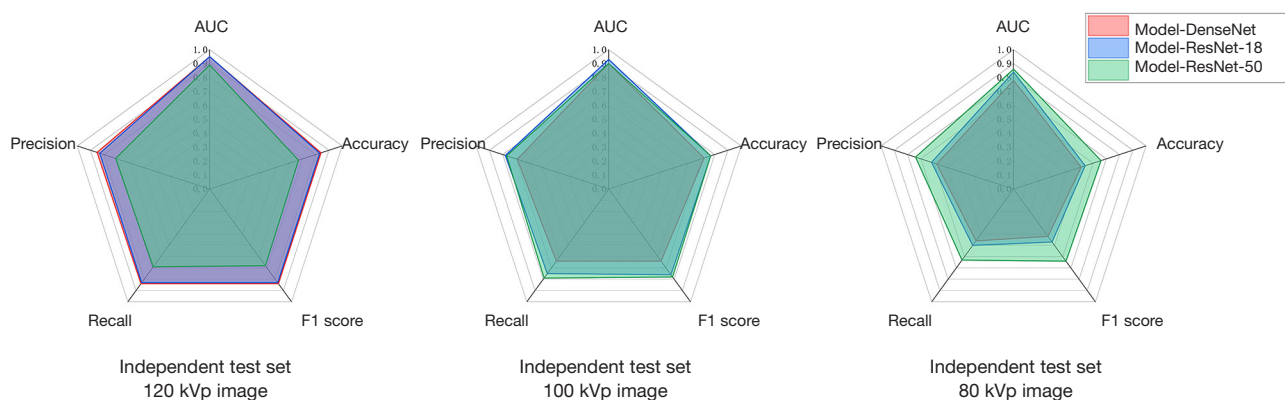


Figure 5 The radar plots were used to display the comprehensive diagnostic performance of the three models on the independent test set. The diagnostic performance of the deep learning models based on DenseNet, ResNet-18, and ResNet-50 shows varying degrees of decline on the independent test set from 120 to 80 kVp images. AUC, area under the curve.

markedly mitigates the time- and labor-intensive aspects of segmentation, contributing to heightened efficiency and accuracy. Among various CNN architectures, the developed ResNet-18 network emerged as the most proficient in bone density status assessment, attaining AUCs of 0.97, 0.91, and 0.95 for detecting normal BMD, osteopenia, and osteoporosis, respectively.

Localization and segmentation of the vertebral body constitute fundamental steps in assessing BMD in chest CT images. Although manual delineation by radiologists is feasible, it demands considerable time, meticulous attention, and consistency. Given the substantial volume of chest CT scans conducted annually, this would pose a substantial workload for clinical practitioners. The advent of DL provides a potential solution to this challenge. Chen *et al.* achieved automatic segmentation of the thoracic spine with a CNN, yielding a Dice coefficient exceeding 0.85 (22). Niu *et al.* successfully localized T12–L2 targets in CT scans

using a DL system (16). However, the standard chest CT scanning range typically excludes the first and second lumbar vertebrae, despite including the lung apices to the bilateral rib diaphragm angles. Some studies have explored the analysis of the entire thoracic spine, but the uneven distribution of BMD in this region may diminish sensitivity in osteoporosis detection (23). Rühling *et al.* observed a gradual increase in the correlation between C2–T12 and lumbar vertebrae (L1–L2) (range, $r_{C2} = 0.76$ to $r_{T12} = 0.96$) (24). Budoff *et al.* proposed that the cancellous bone of the lower thoracic vertebrae (T10–T12) provides pertinent information for BMD assessment (25). The strength of our integrated DL framework lies in its capability to swiftly locate specific T10–T12 regions and precisely delineate TVCB boundaries through a multi-scale strategy. With a clear demarcation between cortex and cancellous regions, the segmentation model exhibited satisfactory performance and accurate segmentation across test sets acquired at various tube

voltages. In 80, 100, and 120 kVp images, the mean Dice coefficient of the segmentation model surpassed 0.93.

The DL networks employed in the model classification phase automatically learn complex features from the input image, enabling end-to-end classification without the need for manually designed hard-coded feature extraction (26). These DL networks consist of various CNNs variants, such as ResNet-18, ResNet-50, and DenseNet. ResNet tackles the challenges of gradient explosion and vanishing problems in CNNs by utilizing multiple stacked ResNet units (27). ResNet-50, characterized by a bottleneck structure for each ResNet block and a deeper architecture than ResNet-18, has demonstrated superior classification accuracy in studies focusing on brain abnormalities and oral squamous cell carcinoma with multiple classes (28-30). Although ResNet-50, with its 50-layer depth, exhibits enhanced expressive power for capturing complex features, beneficial for discriminating tumors with high heterogeneity, it did not perform as well as ResNet-18 in our study for osteoporosis detection, achieving only 0.67 accuracy.

We attribute this discrepancy to the fact that ResNet-50's increased depth and complexity may be counterproductive for a simpler classification task such as osteoporosis detection. This observation aligns with findings by Lu *et al.* in a binary task of identifying primary and metastatic brain tumors (31). ResNet-18, producing more modest features and being lighter than ResNet-50, is less prone to overfitting. Previous studies have demonstrated the effectiveness of ResNet-18 in detecting osteoporosis in lumbar spine X-ray radiographs, achieving an AUC of 0.8 (32). In our study, the ResNet-18 network achieved an AUC exceeding 0.9 for osteoporosis detection in both the internal and independent test sets, benefiting from a larger sample size and more voxels in the chest CT images. DenseNet, a CNN with dense connections between any 2 layers, enhances the transfer and utilization of features while mitigating gradient vanishing during training (31). In our study, DenseNet exhibited notable performance in detecting osteoporosis, with an AUC of 0.95, comparable to the efficacy reported by Niu *et al.* in detecting osteoporosis on T12–L2 (16).

Osteopenia diagnosis involves 2 threshold values: 80 and 120 mg/cm³. Compared to normal BMD and osteoporosis, which have a single threshold value, osteopenia cases are more susceptible to approaching the threshold and prone to classification errors. Thus, diagnosing osteopenia represents a challenging and crucial aspect of the classification model (32).

Both the DenseNet and ResNet-18 networks achieved an AUC exceeding 0.90, indicating their effectiveness in detecting osteopenia. This study separately assessed the male and female populations, considering variations in hormone levels and physical activity. The findings suggest that the performance of the DenseNet and ResNet-18 models remains unaffected by gender changes. Pan *et al.* (33) also utilized a DenseNet-based model to assess BMD from CT images, reporting AUCs of 0.875 and 0.950 for diagnosing osteoporosis in male and female test sets, respectively, which were similar to our study, further validating the utility of our model.

This study is the first to assess the impact of tube voltage on DL classification model performance. Utilizing the DenseNet and ResNet-18 networks, which exhibit superior overall performance, on an independent test set containing images with varying tube voltages, we observed that reducing the tube voltage from 120 to 100 kVp resulted in an average decrease in accuracy and precision of 16.56% for the DenseNet model and 7.74% for the ResNet-18 model. Similarly, decreasing the tube voltage from 120 to 80 kVp led to an average decrease in accuracy and precision of approximately 35.53% for the DenseNet model and 30.95% for the ResNet-18 model. In essence, as the tube voltage decreased, the DL model's ability to assess bone density status diminished, with the DenseNet network being more affected compared to ResNet-18. Consequently, taking all factors into account, we chose ResNet-18 as the backbone network for the classification phase.

This study has certain limitations that warrant attention. Firstly, validation of these findings is necessary with a multicenter dataset and a larger patient population. Secondly, given the rapid development of CNN models, our analysis covered only a limited number of models for diagnostic accuracy. It is imperative to validate the performance of other CNN models. Lastly, this study concentrated on spine bone density measurements, yet the relationship between hip bone density measurements and overall bone density remains unexplored.

Conclusions

We have introduced a cutting-edge DL framework model that offers an effective and efficient strategy for CT-based opportunistic osteoporosis screening without incurring additional costs or radiation exposure. The fully automated nature and high accuracy of the method represent a significant step forward in developing more efficient systems

to support clinical decision-making.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1617/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1617/coif>). J.C. is an employee of Shanghai United Imaging Intelligence (2022–2023), which developed the uAI Research Portal V1.1 used in this study. J.L. serves as the Chief Scientist at GE Healthcare, which produced the CT scanner employed in this research, during the study period of 2022–2023. The two authors were not involved in the data collection or processing phase, ensuring no potential impact on the study's outcomes. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the First Affiliated Hospital of Dalian Medical University (No. PJ-KS-KY-2023-276) and individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Sözen T, Özışık L, Başaran NÇ. An overview and management of osteoporosis. *Eur J Rheumatol* 2017;4:46-56.
- Borgström F, Karlsson L, Ortsäter G, Norton N, Halbout P, Cooper C, Lorentzon M, McCloskey EV, Harvey NC, Javaid MK, Kanis JA; International Osteoporosis Foundation. Fragility fractures in Europe: burden, management and opportunities. *Arch Osteoporos* 2020;15:59.
- Si L, Winzenberg TM, Jiang Q, Chen M, Palmer AJ. Projection of osteoporosis-related fractures and costs in China: 2010–2050. *Osteoporos Int* 2015;26:1929-37.
- Pisani P, Renna MD, Conversano F, Casciaro E, Di Paola M, Quarta E, Muratore M, Casciaro S. Major osteoporotic fragility fractures: Risk factor updates and societal impact. *World J Orthop* 2016;7:171-81.
- Cheng X, Zhao K, Zha X, Du X, Li Y, Chen S, et al. Opportunistic Screening Using Low-Dose CT and the Prevalence of Osteoporosis in China: A Nationwide, Multicenter Study. *J Bone Miner Res* 2021;36:427-35.
- Dimai HP. Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-score, and reference databases. *Bone* 2017;104:39-43.
- Tong X, Fang X, Wang S, Fan Y, Wei W, Xiao Q, Chen A, Liu Y, Liu L. Virtual unenhanced images derived from dual-energy computed tomography for assessing bone mineral density and detecting osteoporosis. *Quant Imaging Med Surg* 2023;13:6571-82.
- Gillespie CW, Morin PE. Trends and Disparities in Osteoporosis Screening Among Women in the United States, 2008-2014. *Am J Med* 2017;130:306-16.
- McAdam-Marx C, Unni S, Ye X, Nelson S, Nickman NA. Effect of Medicare reimbursement reduction for imaging services on osteoporosis screening rates. *J Am Geriatr Soc* 2012;60:511-6.
- Berrington de González A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, Land C. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Arch Intern Med* 2009;169:2071-7.
- Gausden EB, Nwachukwu BU, Schreiber JJ, Lorch DG, Lane JM. Opportunistic Use of CT Imaging for Osteoporosis Screening and Bone Density Assessment: A Qualitative Systematic Review. *J Bone Joint Surg Am* 2017;99:1580-90.
- He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on image classification. 2015 IEEE International

- Conference on Computer Vision (ICCV), Santiago, Chile, 2015:1026-34.
13. 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* 2022;37:369-77.
 14. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. *Eur Radiol* 2020;30:3549-57.
 15. Peng T, Zeng X, Li Y, Li M, Pu B, Zhi B, Wang Y, Qu H. A study on whether deep learning models based on CT images for bone density classification and prediction can be used for opportunistic osteoporosis screening. *Osteoporos Int* 2024;35:117-28.
 16. Niu X, Huang Y, Li X, Yan W, Lu X, Jia X, Li J, Hu J, Sun T, Jing W, Guo J. 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* 2023;13:5294-305.
 17. American College of Radiology. ACR-SPR-SSR practice parameter for the performance of musculoskeletal quantitative computed tomography (QCT). Available online: <https://www.acr.org/-/media/ACR/Files/Practice-Parameters/qct.pdf?la=en>. Accessed 23 Feb 2021.
 18. Dong H, Yin L, Chen L, Wang Q, Pan X, Li Y, Ye X, Zeng M. Establishment and validation of a radiological-radiomics model for predicting high-grade patterns of lung adenocarcinoma less than or equal to 3 cm. *Front Oncol* 2022;12:964322.
 19. Zhu W, Huang H, Zhou Y, Shi F, Shen H, Chen R, Hua R, Wang W, Xu S, Luo X. Automatic segmentation of white matter hyperintensities in routine clinical brain MRI by 2D VB-Net: A large-scale study. *Front Aging Neurosci* 2022;14:915009.
 20. Ma CY, Zhou JY, Xu XT, Guo J, Han MF, Gao YZ, Du H, Stahl JN, Maltz JS. Deep learning-based auto-segmentation of clinical target volumes for radiotherapy treatment of cervical cancer. *J Appl Clin Med Phys* 2022;23:e13470.
 21. Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira SP, Clarkson MJ, Barratt DC. Automatic Multi-Organ Segmentation on Abdominal CT With Dense V-Networks. *IEEE Trans Med Imaging* 2018;37:1822-34.
 22. Chen YC, Li YT, Kuo PC, Cheng SJ, Chung YH, Kuo DP, Chen CY. Automatic segmentation and radiomic texture analysis for osteoporosis screening using chest low-dose computed tomography. *Eur Radiol* 2023;33:5097-106.
 23. Naghavi M, De Oliveira I, Mao SS, Jaberzadeh A, Montoya J, Zhang C, Atlas K, Manubolu V, Montes M, Li D, Atlas T, Reeves A, Henschke C, Yankelevitz D, Budoff M. Opportunistic AI-enabled automated bone mineral density measurements in lung cancer screening and coronary calcium scoring CT scans are equivalent. *Eur J Radiol Open* 2023;10:100492.
 24. Rühling S, Scharr A, Sollmann N, Wostrack M, Löffler MT, Menze B, Sekuboyina A, El Husseini M, Braren R, Zimmer C, Kirschke JS. Proposed diagnostic volumetric bone mineral density thresholds for osteoporosis and osteopenia at the cervicothoracic spine in correlation to the lumbar spine. *Eur Radiol* 2022;32:6207-14.
 25. Budoff MJ, Hamirani YS, Gao YL, Ismaeel H, Flores FR, Child J, Carson S, Nee JN, Mao S. Measurement of thoracic bone mineral density with quantitative CT. *Radiology* 2010;257:434-40.
 26. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
 27. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Martinez-Gonzalez P, Garcia-Rodriguez J. A survey on deep learning techniques for image and video semantic segmentation. *Appl Soft Comput* 2018;70:41-65.
 28. Talo M, Yildirim O, Baloglu UB, Aydin G, Acharya UR. Convolutional neural networks for multi-class brain disease detection using MRI images. *Comput Med Imaging Graph* 2019;78:101673.
 29. Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Netw* 2020;128:47-60.
 30. He K, Zhang X, Ren S, Sun J. Deep ResNet learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:770-8.
 31. Lu SY, Satapathy SC, Wang SH, Zhang YD. PBTNet: A New Computer-Aided Diagnosis System for Detecting Primary Brain Tumors. *Front Cell Dev Biol* 2021;9:765654.
 32. Zhang B, Chen Z, Yan R, Lai B, Wu G, You J, Wu X,

Duan J, Zhang S. Development and Validation of a Feature-Based Broad-Learning System for Opportunistic Osteoporosis Screening Using Lumbar Spine Radiographs. *Acad Radiol* 2024;31:84-92.

33. Pan Y, Shi D, Wang H, Chen T, Cui D, Cheng X, Lu Y. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. *Eur Radiol* 2020;30:4107-16.

Cite this article as: Wang S, Tong X, Cheng Q, Xiao Q, Cui J, Li J, Liu Y, Fang X. Fully automated deep learning system for osteoporosis screening using chest computed tomography images. *Quant Imaging Med Surg* 2024;14(4):2816-2827. doi: 10.21037/qims-23-1617