EUROPEAN SOCIETY OF RADIOLOGY

European Radiology

Check for updates

MUSCULOSKELETAL

Artificial intelligence assisted automatic screening of opportunistic osteoporosis in computed tomography images from different scanners

Yan Wu[1], Xiaopeng Yang[1], Mingyue Wang[1], Yanbang Lian[1], Ping Hou[1], Xiangfei Chai[2], Qiong Dai[2], Baoxin Qian[2], Yaojun Jiang[1]* and Jianbo Gao[1]*

Abstract

Objectives It is feasible to evaluate bone mineral density (BMD) and detect osteoporosis through an artificial intelligence (AI)-assisted system by using quantitative computed tomography (QCT) as a reference without additional radiation exposure or cost.

Methods A deep-learning model developed based on 3312 low-dose chest computed tomography (LDCT) scans (trained with 2337 and tested with 975) achieved a mean dice similarity coefficient of 95.8% for T1–T12, L1, and L2 vertebral body (VB) segmentation on test data. We performed a model evaluation based on 4401 LDCT scans (obtained from scanners of 3 different manufacturers as external validation data). The BMD values of all individuals were extracted from three consecutive VBs: T12 to L2. Line regression and Bland–Altman analyses were used to evaluate the overall detection performance. Sensitivity and specificity were used to evaluate the diagnostic performance for normal, osteopenia, and osteoporosis patients.

Results Compared with the QCT results as the diagnostic standard, the BMD assessed had a mean error of $(-0.28, 2.37)$ mg/cm^3. Overall, the sensitivity of a normal diagnosis was greater than that of a diagnosis of osteopenia or osteoporosis. For the diagnosis of osteoporosis, the model achieved a sensitivity > 86% and a specificity > 98%.

Conclusion The developed tool is clinically applicable and helpful for the positioning and analysis of VBs, the measurement of BMD, and the screening of osteopenia and osteoporosis.

Clinical relevance statement The developed system achieved high accuracy for automatic opportunistic osteoporosis screening using low-dose chest CT scans and performed well on CT images collected from different scanners.

Key Points

- *Osteoporosis is a prevalent but underdiagnosed condition that can increase the risk of fractures.*
- *This system could automatically and opportunistically screen for osteoporosis using low-dose chest CT scans obtained for lung cancer screening.*
- *The developed system performed well on CT images collected from different scanners and did not differ with patient age or sex.*

*Correspondence:
Yaojun Jiang
jiangyaojun_3526@163.com
Jianbo Gao
cjr.gaojianbo@vip.163.com
[1]Department of Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China
[2]Department of Scientific Research, Huiying Medical Technology, Beijing, China

Springer

Introduction

Osteoporosis is recognized as a latent metabolic bone disease characterized by low bone mass and micro-architectural deterioration resulting in increased bone fragility [1–4]. Osteoporosis is underdiagnosed. It remains asymptomatic for several years until the development of fractures that limit daily life activities, especially for elderly people [5–7]. As the population ages, the number of patients in China with osteoporosis or osteoporotic fractures is estimated to be 212 million and 5.99 million, respectively, by 2050, which will lead to significant increases in morbidity and mortality [8, 9]. Therefore, early screening and monitoring of osteoporosis are crucial for timely prevention and treatment of osteoporotic fractures [10–12]. BMD is an important marker of bone strength. Osteoporosis is diagnosed when the BMD falls below a certain threshold and can also predict fracture risk [1, 13–15]. Although dual-energy X-ray absorptiometry (DXA)-derived areal BMD is required for diagnosing osteoporosis according to the World Health Organization criteria, Quantitative CT (QCT) is more sensitive for monitoring bone loss and measuring trabecular bone density because it is less affected by severe degeneration of the spine, vascular calcification, oral contrast agents, or body position [16–18]. Low-dose chest computed tomography (LDCT) scans performed for lung cancer screening, including lumbar spine images, can be used to assess volumetric bone mineral density and screen for osteoporosis simultaneously [19, 20].

Although QCT has been proposed as a relatively accurate method for evaluating BMD, it requires appropriate equipment to be deployed, which is expensive for many, especially basic-level hospitals, to afford. In previous studies [14, 21], conventional diagnostic CT scans were used to measure BMD via direct measurement of the CT values of cancellous bone, with correlation coefficients ranging from 0.399 to 0.891. However, CT image analysis requires frequent manual operations, including localization of vertebral bodies (VBs) and extraction of the volume of interest (VOI), which are heavy and redundant tasks in large-scale osteoporosis screening [6, 22]. With the current advancements in computational performance, data processing, and the availability of large-scale datasets, artificial intelligence (AI)-based medical image analysis has played an essential role in extracting image features, identifying lesion areas, and classifying disease types, which provides helpful assistance in the diagnosis and prevention of disease [23]. In particular, deep-learning (DL) algorithms, especially convolutional neural network (CNN) algorithms, have been applied to extract vertebral CT values to determine BMD and related fractures from CT images [11, 24]. DL is expected to replace manual operation in BMD measurement, to liberate radiologists for more important and meaningful tasks, and to reduce the cost of osteoporosis screening [25, 26].

There have been a few studies reporting the application of DL techniques for osteoporosis screening. Pan Yali et al [27] developed a DL-based system to automatically measure BMD for opportunistic osteoporosis screening from low-dose chest CT scans for lung cancer screening. However, in this report, only images from a single CT scanner were utilized for training and testing the DL-based model. Yasaka et al [28] developed a DL model to predict the BMD of lumbar vertebrae from unenhanced abdominal CT images by using BMD levels obtained from DXA as a reference. The model derived from this study focused on the prediction of a real BMD but did not provide information on VB location. Fang Yijie et al [29] developed a deep CNN model to predict the BMD of lumbar vertebrae (L1–L4) with BMD values obtained from QCT as a reference. LDCT for lung cancer screening, which causes neither additional exposure nor substantial cost, is more common for spinal or abdominal CT scans. Li Yali et al [30, 31] demonstrated that the qualities of LDCT images obtained from various CT scanners sold by five different manufacturers were comparable. The linearity of CT values from those LDCT images was unbiased and could be utilized to quantify the BMD values accurately. However, few studies in the literature have focused on the validation of DL-based automatic osteoporosis screening models in LDCT images obtained from different scanners because of their accuracy in locating the thoracic and upper lumbar regions and in calculating BMD values, which could be comparable to those obtained from QCT.

In this study, we aimed to (I) identify the potential of CT scans, which were originally used for lung cancer screening, for both osteoporosis screening and bone quality evaluation and (II) assess the performance of the AI-based automatic osteoporosis screening model in both locating the lumbar VB and calculating BMD values in CT images obtained from different CT scanners.

Materials and methods
Study design
This retrospective, single-center study was approved by the First Affiliated Hospital of Zhengzhou University's Institutional Review Board (No. 2021-KY-1222-002), and the need for informed consent was waived. From the electronic database of our hospital, we retrieved the data

of individuals who underwent paired LDCT and QCT examinations for screening both lung cancer and osteoporosis from September 2019 to January 2022. The collected data were divided into two parts, and one part was manually annotated to obtain the diagnostic standard for segmenting vertebrae. It is used for training and testing to explore the effect of different gender and age groups on the model. The other part was used for external validation to explore the effect of scanners from different manufacturers on the model. The development of a fully automated BMD measurement system consisted of two main steps. First, an end-to-end DL model was trained to achieve automatic segmentation from the first thoracic vertebra to the second lumbar vertebra (T1–T12, L1, and L2). Second, the DL regression detection algorithm was used to calculate the BMD from the twelfth thoracic vertebra to the second lumbar vertebra (T12, L1, and L2). The workflow is shown in Fig. 1.

Dataset

Individuals who had a history of prior spinal surgery or primary or metastatic tumors were excluded. Finally, 7713 individuals were enrolled in this study. A total of 3312 LDCT scans were manually annotated by an experienced radiologist for the contours and the anatomical names of all VBs and were used to develop the DL-based system. These images were randomly divided into a training set ($N = 2337$) and a test set ($N = 975$). The categorization process ensured that the gender distribution was balanced

and that the age distribution was realistic. The remaining 4401 unannotated LDCT scans obtained from 3 different scanners (200 scans from GE, 2867 scans from Philips, and 1334 scans from Siemens) were used as external validation datasets to evaluate the developed system.

Data preparation
Data collection
We chose hospitals with QCT scanners from different vendors to collect the data. QCT can be carried out simultaneously with LDCT during plain-scan physical examination. Both the low-dose chest CT images and the corresponding BMD values could be collected via QCT. In accordance with previous studies, all individuals were scanned from the apical lung to the lower edge of L2 on three different scanners at the same hospital ((I) Revolution CT, GE Healthcare; (II) Brilliance iCT, Philips Health care; (III) Somatom Force, Siemens Healthineers). Non-contrast chest CT was performed while the patient held their mid-inspiratory breath. To calibrate the linear relationship between CT values and BMD values, QA phantom data were acquired once a month by separately scanning an asynchronous phantom (Mindways Software, Inc.) with the same parameters as other scans. The scan parameters are shown in Supplemental Table 1 [19]. BMD values were recorded directly by the physician operating the QCT equipment. The BMD values of all individuals were extracted from three consecutive VBs, T12 to L2, using QCT image analysis software (QCT Pro 6.1,
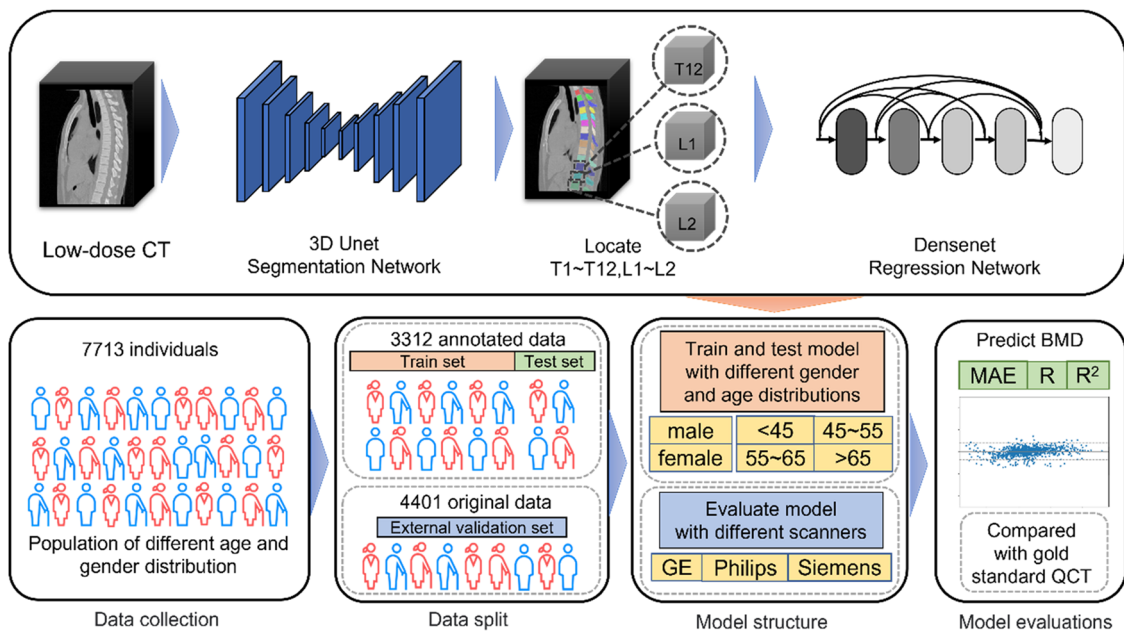

Fig. 1 Overview of this study

Mindways Software, Inc.) [27]. The 9 mm-high VOIs capturing the trabecular bone were manually placed in the center of the targeted VBs, avoiding the basivertebral veins, cortical bone, and any focal pathology [19]. Based on the standard clinically utilized average BMD of L1–L2, bone mass was categorized as normal ($> 120 \, \text{mg/cm}^3$), osteopenia (80–$120 \, \text{mg/cm}^3$), or osteoporosis ($< 80 \, \text{mg/cm}^3$) [32]. The BMD values from T12 to L2 and the assessment of bone mass constitute the final output of the developed system.

Data annotation

The diagnostic standard of labeling specifications was developed by three imaging physicians, two of whom were annotators and one of whom was an arbitrator. The annotator was a practising physician who had been practising medical imaging diagnosis at a public medical institution for more than 5 years, and the arbitrator was a practising physician who had been practising medical imaging diagnosis at a public medical institution for more than 10 years [29]. The diagnostic standard was developed as follows: two annotating physicians independently reviewed the images, marked the target spine region in their own images, and then passed their images and annotation results to the arbitrator physician separately. The arbitrator physician-reviewed and modified the results of the two annotating physicians to form a unique final annotation result.

Statistical analysis

Continuous variables with an abnormal distribution are expressed as medians (IQRs), and categorical variables are represented as frequencies. The chi-square test was carried out to assess categorical variables. The Kruskal–Wallis test was used to compare abnormally distributed continuous variables between the two datasets. The Pearson correlation coefficient R was used in the correlation evaluation for the bone density regression task. We used the mean absolute error (MAE), coefficient of determination (R^2), and Bland–Altman plot to evaluate the overall regression detection performance. The Kolmogorov–Smirnov test was used to test the normality of all the continuous variables. $p < 0.05$ was considered to indicate statistical significance. R software for Windows (version 4.2.1) was used for the statistical analyses. Sensitivity and specificity were used to evaluate the diagnostic performance for normal, osteopenia, and osteoporosis patients.

Results

Table 1 summarizes the demographic characteristics of all patients ($n = 3312$) who were divided into a training set ($n = 2337$) and a test set ($n = 975$). For all patients, the median age was 54 years (IQR, 47–61 years), 1540 were female (46.5%), and 1772 were male (53.5%). We then divided the patients into four groups based on age distribution: under 45 years of age, between 45 and 55 years of age, between 55 and 65 years of age, and over 65 years of age, with 553 (16.7%), 1224 (37.6%), 989 (29.9%) and 526 (15.9%) patients, respectively. The median slice number of each CT scan was 370 (IQR, 345–405). Age and slice number exhibited statistically significant differences between the different groups ($p < 0.05$).

The segmentation results of the VBs for the test set are shown below in Supplemental Fig. S2. The proposed method achieved a mean Dice similarity coefficient (DSC) of 95.8% for VB segmentation. The DSCs of the T12, L1, and L2 vertebrae, which were the focus of the subsequent BMD regression network, were 95.2%, 94.1%, and 95.6%, respectively. Figure 2 shows the original images and the AI automatic segmentation images of VBs from normal, osteopenia, and osteoporosis patients.

Table 1 Baseline patient characteristics

|		Overall (N = 3312)	Train set (N = 2337)	Test set (N = 975)	p-value
Age	Median (IQR)	54 (47, 61)	53 (47, 60)	55 (49, 61)	< 0.0001
Age group (%)	≤ 45	553 (16.7)	434 (18.6)	119 (12.2)	< 0.0001
	45–55	1244 (37.6)	901 (38.6)	343 (35.2)	
	55–65	989 (29.9)	646 (27.6)	343 (35.2)	
	> 65	526 (15.9)	356 (15.2)	170 (17.4)	
Sex (%)	Female	1540 (46.5)	1090 (46.6)	450 (46.2)	0.8275
	Male	1772 (53.5)	1247 (53.4)	525 (53.8)	
BMI	Median (IQR)	24.50 (22.30, 26.70)	24.20 (22.13, 26.37)	24.50 (22.40, 26.80)	0.172
Slicer number	Median (IQR)	370 (345, 405)	370 (350, 463)	370 (345, 390)	0.0004
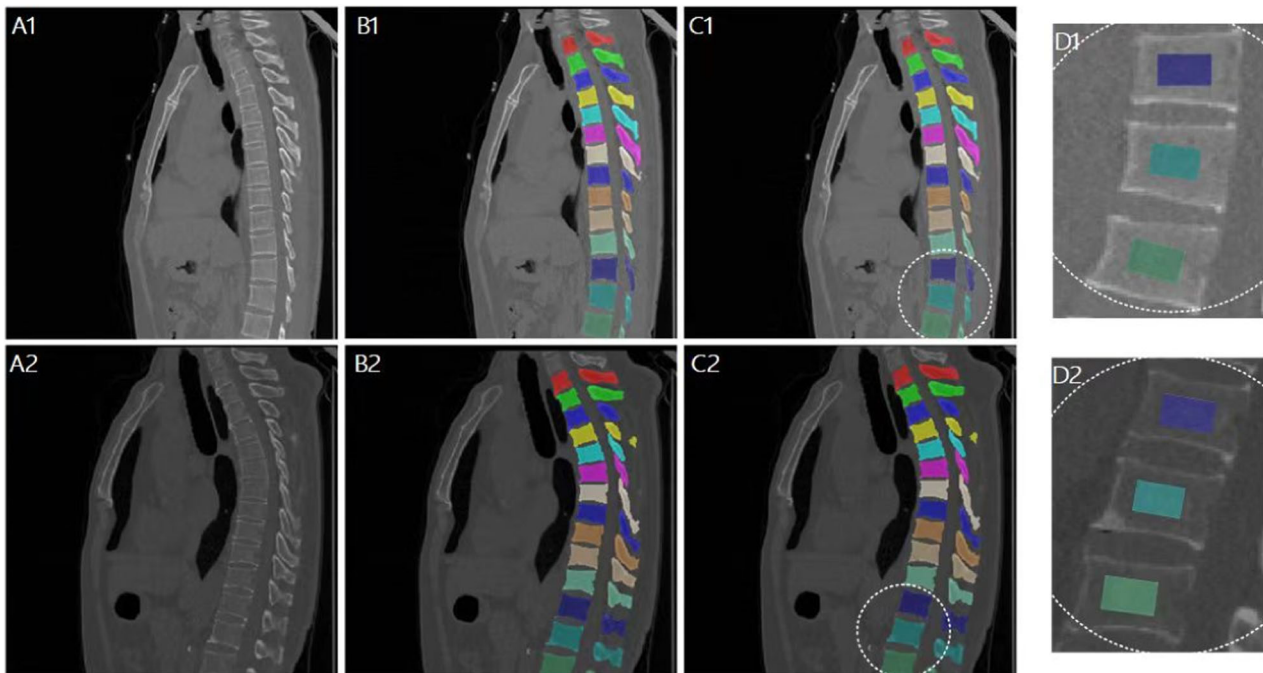
Fig. 2 Segmentation masks with three categories were predicted by the AI model and visualized by ITK-SNAP (**A1**, **A2**) CT sagittal image (**B1**, **B2**) manual segmentation (**C1**, **C2**) automated segmentation (**D1**, **D2**) automated VOI segmentation from T12–L2

Table 2 Overall segmentation results in different gender and age groups on the test set

Type	Distribution	Overall DSC (%)
Gender	Female	95.5
	Male	96.0
Age group	≤ 45	96.1
	45–55	95.9
	55–65	95.5
	Over 65	95.0

Table 3 Comparison of algorithm performance between different genders and age groups on test set

Type	Distribution	MAE	R	R^2
Test set		2.88	0.994	0.987
Gender	Female	3.00	0.995	0.989
	Male	2.75	0.993	0.985
Age group	≤ 45	2.29	0.995	0.989
	45–55	2.63	0.993	0.986
	55–65	2.85	0.987	0.973
	Over 65	3.67	0.988	0.975

The experiments also evaluated the segmentation performance of the model among different sex and age groups, as shown in Table 2. For all VB segmentations obtained from scanners from different manufacturers, R was greater than 0.99, R^2 was greater than 0.98, and the MAE was less than 3.5. The test results showed that the overall VB segmentation performance was better than the segmentation performance of a specific VB. The performance evaluation metrics of the segmentation model in the male and female groups were very close to each other; therefore, gender did not have a significant influence on the segmentation network performance. Similarly, there were no significant differences among patients of different ages for the performance evaluation metrics of the segmentation model, indicating that

patient age did not significantly affect segmentation network performance.

The evaluation metrics indicating the overall performance of the BMD regression algorithm model for different sex and age groups in the test set are shown in Table 3. The Bland−Altman plot of the test set is shown in Fig. 3A, in which the abscissa is the mean value of the predicted and actual BMD. The ordinate is the difference between the predicted BMD and the actual value, and the two dashed lines are the 95% consistency limits. The vast majority of the differences are within this range. When the algorithm model was applied to the validation dataset, the overall correlation coefficient R was 0.994, the determination coefficient R^2 was 0.987, and the MAE was 2.88. To study the influence of the age of different patients on
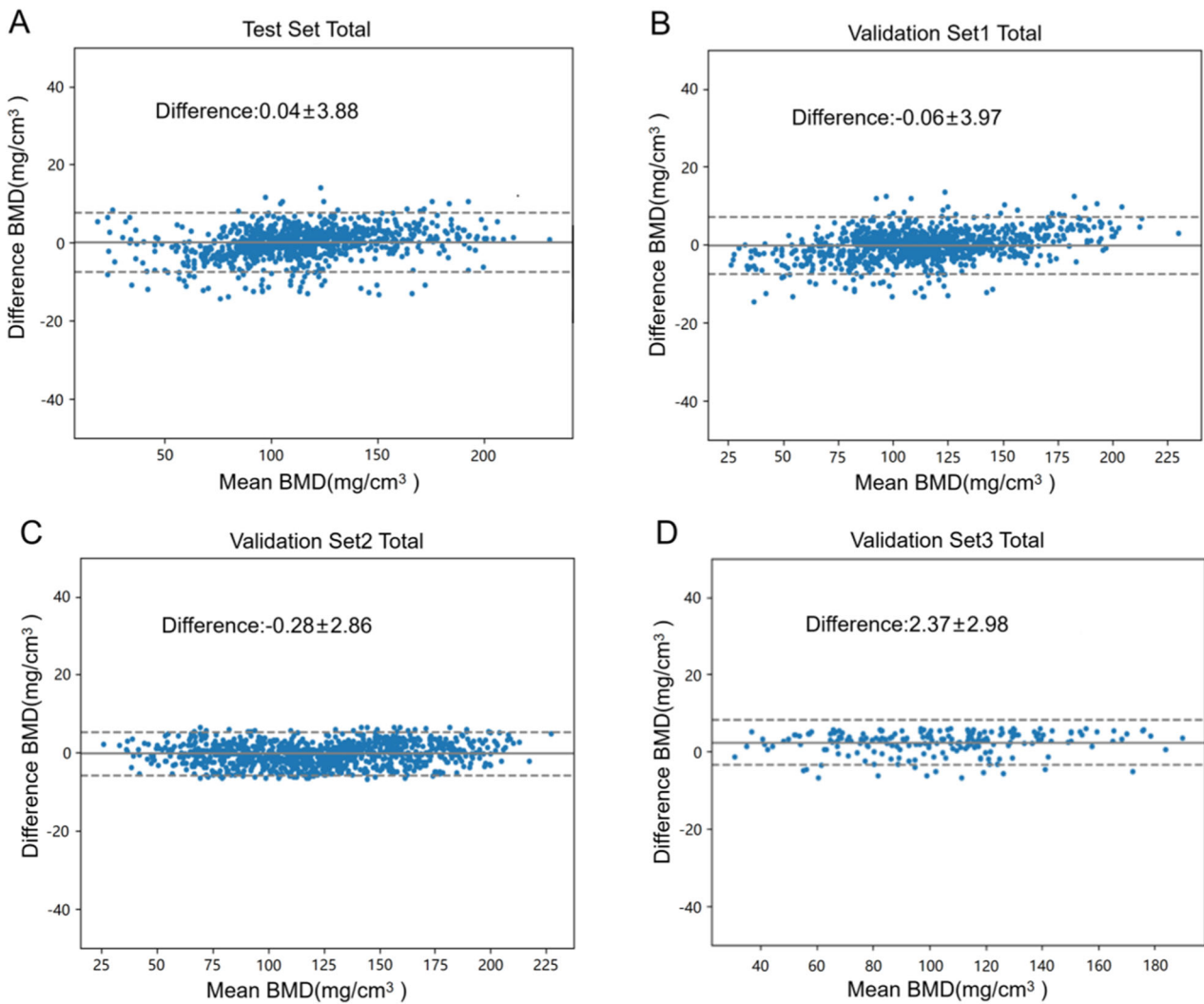
Fig. 3 Bland–Altman plot comparing BMD values obtained by using the developed system and QCT on the test set (**A**) and on external validation sets obtained from scanners of different manufacturers (**B–D**)

the performance of the algorithm model, the patients were divided into four groups based on their ages: under 45 years, 45–55 years, 55–65 years, and over 65 years. There were no significant differences in algorithm performance among the different sex and age groups.

There were 4401 patients in the external validation datasets, including 2867 patients from Philips manufacturers (Set1), 1334 patients from Siemens manufacturers (Set2), and 200 patients from general manufacturers (Set3). All the data were applied to the previous segmentation network to obtain the vertebral masks corresponding to T1–T12 and L1–L2. Linear analysis of BMD values obtained from QCT and algorithmic models based on T12, L1, and L2 vertebral images from different manufacturers' devices. The statistical results are shown in Table 4. A Bland–Altman plot comparing the BMD values obtained by using the developed system and QCT with those obtained by scanners from different manufacturers is shown in Fig. 3B–D. Most of the differences were within this range. Compared with the QCT results as the diagnostic standard, the BMD assessed by using the developed algorithm model had a mean error of $(-0.28, 2.37)$ mg/cm^3, with the 95% consistency limits of Set1, Set2, and Set3 falling in the ranges of $(-3.79, 3.67)$ mg/cm^3, $(-3.14, 2.58)$ mg/cm^3 and $(-0.61, 5.35)$ mg/cm^3, respectively. Based on the results above, it could be concluded that the BMD regression model exhibited an accurate prediction, good generalization ability, and robustness for CT scans obtained from all three different scanners supplied by Philips, Siemens, and general equipment manufacturers.

Table 4 Linear analysis of BMD values obtained from QCT and from the algorithm model based on data collected from scanners of different manufacturers on external validation sets

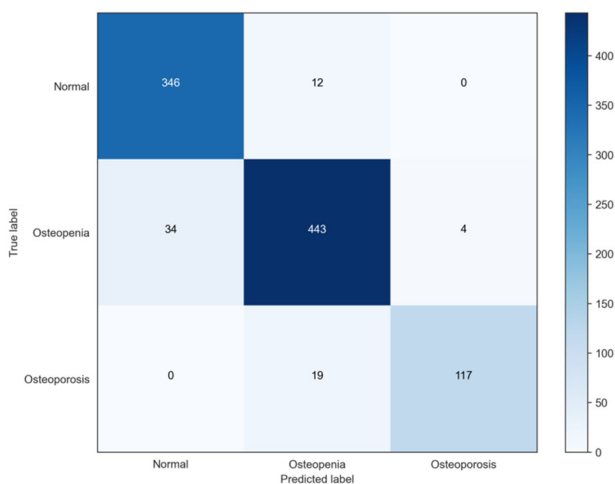| Different manufacturers	VBs	MAE	R	R^2
Set1	T12	4.45	0.989	0.977
	L1	4.31	0.987	0.974
	L2	4.22	0.988	0.976
	Overall	3.41	0.994	0.987
Set2	T12	3.35	0.995	0.990
	L1	3.37	0.993	0.985
	L2	3.59	0.992	0.985
	Overall	2.60	0.997	0.993
Set3	T12	5.99	0.982	0.955
	L1	4.47	0.988	0.971
	L2	3.95	0.991	0.980
	Overall	3.43	0.996	0.986

Table 5 The diagnostic performance of normal, osteopenia, and osteoporosis on test set and external validation sets

Dataset	Sensitivity	Specificity	PPV	NPV
Test set ($N = 975$)				
Normal	96.65%	94.49%	91.05%	97.98%
Osteopenia	92.10%	93.72%	93.46%	92.42%
Osteoporosis	86.03%	99.52%	96.69%	97.78%
Validation Set1 ($N = 2867$)				
Normal	98.17%	92.65%	91.89%	98.36%
Osteopenia	89.32%	95.51%	93.04%	93.02%
Osteoporosis	86.75%	99.64%	97.47%	97.89%
Validation Set2 ($N = 1334$)				
Normal	99.41%	94.20%	94.67%	99.36%
Osteopenia	90.62%	96.61%	93.12%	95.32%
Osteoporosis	87.44%	99.65%	97.84%	97.74%
Validation Set3 ($N = 200$)				
Normal	96.15%	97.97%	94.34%	98.64%
Osteopenia	93.41%	91.74%	90.43%	94.34%
Osteoporosis	87.72%	97.90%	94.34%	95.24%

validation sets, as shown in Table 5. Overall, the sensitivity of a normal diagnosis was greater than that of a diagnosis of osteopenia or osteoporosis. The sensitivity of the diagnosis of osteoporosis was greater in Set3 than in the other two datasets, possibly due to the smaller sample size of this test set compared to the other two datasets. For the diagnosis of osteoporosis, the model achieved a sensitivity of more than 86% and a specificity of more than 98% for different datasets, which could indicate excellent performance.

Discussion

We proposed a model based on DL for fully automated prediction of BMD from CT images with the bone density values acquired from QCT as the gold standard. The model could extract the T12, L1, and L2 vertebrae via direct segmentation of VBs at the 3D level with the 3D-UNet architecture [33]. Subsequently, the modified DenseNet121 network [34] was used for the end-to-end prediction of BMD. We used 3312 cases of data after labeling for the training and testing sets of the model and then analyzed the robustness of the model among patients of different genders and ages. A total of 4401 scans from different CT scanners supplied by three different instrument manufacturers were also collected as external validation sets to evaluate the stability of the algorithm across different devices.

In terms of segmentation performance, our model had an average DCE of 95.8% and showed good stability among patients of different genders and ages. The segmented VB targets in a previous study [29] were the

Fig. 4 Confusion matrix of the test set

We also converted the BMD values predicted by the proposed algorithm model and obtained from QCT into a three-class qualitative classification task for evaluation. A confusion matrix is a cross-tabulation of the model's predicted results and the actual results. The predicted results of the test set are shown in Fig. 4 (a confusion matrix), and the higher the diagonal values are, the better the model performance. However, the confusion matrix only counts the number of cases, and it is difficult to measure the quality of the model when a large amount of data is available. Therefore, the confusion matrix extends the following indicators based on the basic statistical results: sensitivity, specificity, positive predictive value, and negative predictive value. We calculated the above evaluation indicators on the test set and three external

L1–L4 vertebrae, and the average segmentation accuracy was approximately 82.3%. The task was trained based on 2D U-Net only, which missed the spatial features of vertebrae and affected the subsequent analytical results. A similar study [27] segmented T1–T6, T7–T12, and L1–L2 separately based on a 2D network and subsequently segmented the T12, L1, and L2 vertebrae using a conventional image processing algorithm. The average segmentation accuracy was 86.6%. Our approach, different from previous studies, was based on 3D image analysis with an end-to-end network model to obtain results for each individual VB segmentation, achieving much greater accuracy than previous 2D models.

In terms of the accuracy of BMD regression prediction, our model had a high R^2 of 0.987 and an MAE of only 2.88 in the validation set. The MAEs for patients of different sexes ranged from 2.75–3.00, and the MAEs for patients of different age groups ranged from 2.29–3.67, both of which demonstrated a high degree of stability. We also collected CT scans from devices supplied by three different manufacturers for validation of the regression model. The model achieved R^2 values of 0.987, 0.993, and 0.986 and R values of 0.994, 0.992, and 0.996 in Set1, Set2, and Set3, respectively, with MAEs in the range of 2.60–3.41. Compared to previously reported studies [27, 29], our models achieved the best performance.

We also evaluated the diagnostic efficacy of qualitative osteoporosis screening based on the predicted BMD values. The sensitivity of the model was greater than 86% for osteoporosis and greater than 89% for osteopenia among the validation set and different external validation sets. It should be noted that the attribution of BMD values at the discriminant boundary (80 and 120 mg/cm^3) could influence the diagnostic efficacy of the classification. Therefore, comparing the assessed BMD values acquired from the algorithm models and those from the QCT method to evaluate bone quality is more reasonable than attributing patients to only three classifications.

In summary, a clinically useful opportunistic osteoporosis screening tool has been developed based on DL combined with CT images from various scanners of different manufacturers. This tool achieves excellent accuracy for both the 3D segmentation of VBs and the regression prediction of BMD. Therefore, it has a strong ability to screen for osteoporosis with high efficacy and sensitivity. In the near future, the proposed model can be further improved in the following aspects. More special cases of patients with incomplete numbers of vertebrae for various reasons (e.g., injury, surgery, congenital abnormalities, etc.) should be collected and utilized to train the model for further improvement of the precision of VB localization and segmentation for those special patients (approximately 5% of all patients). More data should be collected from different hospitals or medical centers all over China to further validate the applicability, adaptability, and robustness of this proposed osteoporosis screening and bone quality assessment algorithm model.

Abbreviations

AI	Artificial intelligence
BMD	Bone mineral density
CNN	Convolutional neural network
DL	Deep learning
DSC	Dice similarity coefficient
DXA	Dual-energy X-ray absorptiometry
LDCT	Low-dose chest computed tomography
MAE	Mean absolute error
QCT	Quantitative computed tomography
VBS	Vertebral body
VOI	Volume of interest

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1007/s00330-024-11046-2.

Compliance with ethical standards

Guarantor

This study was guaranteed by Yan Wu, Department of Radiology, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China.

Conflict of interest

Yan Wu, Xiaopeng Yang, Mingyue Wang, Yanbang Lian, Ping Hou, Xiangfei Chai, Qiong Dai, Baoxin Qian, Yaojun Jiang, and Jianbo Gao declare that they have no conflicts of interest.

Statistics and biometry

One of the authors has significant statistical expertise.

Informed consent

Written informed consent was waived by the Institutional Review Board.

Ethical approval

Institutional Review Board approval was obtained.

Study subjects or cohorts overlap

No study subjects or cohorts have been previously reported.

Methodology

- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

References

1. NIH Consensus Development Panel on Osteoporosis Prevention D, and Therapy (2001) Osteoporosis prevention, diagnosis, and therapy. JAMA 285:785–795
2. Cummings SR, Melton LJ (2002) Epidemiology and outcomes of osteoporotic fractures. Lancet 359:1761–1767
3. Rachner TD, Khosla S, Hofbauer LC (2011) Osteoporosis: now and the future. Lancet 377:1276–1287
4. Kanis JA, McCloskey EV, Johansson H et al (2013) European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 24:23–57
5. Ensrud KE, Schousboe JT (2011) Clinical practice. Vertebral fractures. N Engl J Med 364:1634–1642
6. Lee JH, Lee YK, Oh SH et al (2016) A systematic review of diagnostic accuracy of vertebral fracture assessment (VFA) in postmenopausal women and elderly men. Osteoporos Int 27:1691–1699
7. Gates M, Pillay J, Nuspl M, Wingert A, Vandermeer B, Hartling L (2023) Screening for the primary prevention of fragility fractures among adults aged 40 years and older in primary care: systematic reviews of the effects and acceptability of screening and treatment, and the accuracy of risk prediction tools. Syst Rev 12:51
8. Si L, Winzenberg TM, Jiang Q, Chen M, Palmer AJ (2015) Projection of osteoporosis-related fractures and costs in China: 2010-2050. Osteoporos Int 26:1929–1937
9. Agten CA, Ramme AJ, Kang S, Honig S, Chang G (2017) Cost-effectiveness of virtual bone strength testing in osteoporosis screening programs for postmenopausal women in the United States. Radiology 285:506–517
10. Liu J, Wang H, Shan X et al (2024) Hybrid transformer convolutional neural network-based radiomics models for osteoporosis screening in routine CT. BMC Medical Imaging 24:62
11. Engelke K, Chaudry O, Bartenschlager S (2023) Opportunistic screening techniques for analysis of CT scans. Curr Osteoporos Rep 21:65–76
12. Grad R, Reynolds DL, Antao V et al (2023) Screening for primary prevention of fragility fractures: How much time does it take? Can Fam Physician 69:537–541
13. Ahern DP, McDonnell JM, Riffault M et al (2021) A meta-analysis of the diagnostic accuracy of Hounsfield units on computed topography relative to dual-energy X-ray absorptiometry for the diagnosis of osteoporosis in the spine surgery population. Spine J 21:1738–1749
14. Kim YW, Kim JH, Yoon SH et al (2017) Vertebral bone attenuation on low-dose chest CT: quantitative volumetric analysis for bone fragility assessment. Osteoporos Int 28:329–338
15. Genant HK, Boyd D (1977) Quantitative bone mineral analysis using dual energy computed tomography. Invest Radiol 12:545–551
16. Michalski AS, Besler BA, Burt LA, Boyd SK (2021) Opportunistic CT screening predicts individuals at risk of major osteoporotic fracture. Osteoporos Int 32:1639–1649
17. Engelke K, Libanati C, Liu Y et al (2009) Quantitative computed tomography (QCT) of the forearm using general purpose spiral whole-body CT scanners: accuracy, precision and comparison with dual-energy X-ray absorptiometry (DXA). Bone 45:110–118
18. Buckens CF, Dijkhuis G, de Keizer B, Verhaar HJ, de Jong PA (2015) Opportunistic screening for osteoporosis on routine computed tomography? An external validation study. Eur Radiol 25:2074–2079
19. Wu Y, Guo Z, Fu X et al (2019) The study protocol for the China Health Big Data (China Biobank) project. Quant Imaging Med Surg 9:1095–1102
20. Salzmann SN, Shirahata T, Yang J et al (2019) Regional bone mineral density differences measured by quantitative computed tomography: does the standard clinically used L1-L2 average correlate with the entire lumbosacral spine? Spine J 19:695–702
21. Gausden EB, Nwachukwu BU, Schreiber JJ, Lorich DG, Lane JM (2017) Opportunistic use of CT imaging for osteoporosis screening and bone density assessment: a qualitative systematic review. J Bone Joint Surg Am 99:1580–1590
22. Smets J, Shevroja E, Hugle T, Leslie WD, Hans D (2021) Machine learning solutions for osteoporosis-A review. J Bone Miner Res 36:833–851
23. Meng J, Sun N, Chen Y et al (2019) Artificial neural network optimizes self-examination of osteoporosis risk in women. J Int Med Res 47:3088–3098
24. Tang C, Zhang W, Li H et al (2021) CNN-based qualitative detection of bone mineral density via diagnostic CT slices for osteoporosis screening. Osteoporos Int 32:971–979
25. Rastegar S, Vaziri M, Qasempour Y et al (2020) Radiomics for classification of bone mineral loss: a machine learning study. Diagn Interv Imaging 101:599–610
26. Pickhardt PJ, Nguyen T, Perez AA et al (2022) Improved CT-based osteoporosis assessment with a fully automated deep learning tool. Radiol Artif Intell 4:e220042
27. Pan Y, Shi D, Wang H et al (2020) Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. Eur Radiol 30:4107–4116
28. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O (2020) Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. Eur Radiol 30:3549–3557
29. Fang Y, Li W, Chen X et al (2021) Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. Eur Radiol 31:1831–1842
30. Li Y, Jiang Y, Liu H et al (2022) A phantom study comparing low-dose CT physical image quality from five different CT scanners. Quant Imaging Med Surg 12:766–780
31. Li Y, Jiang Y, Yu X et al (2022) Deep-learning image reconstruction for image quality evaluation and accurate bone mineral density measurement on quantitative CT: a phantom-patient study. Front Endocrinol 13:884306
32. American College of Radiology (2021) ACR–SPR–SSR practice parameter for the performance of musculoskeletal quantitative computed tomography (QCT). Available at: chrome-extension://bnjoienjhhclcabnkbhhfn-decoipmcdg/background/jgpdf/layout/index.html?file=https://www.acr.org/-/media/ACR/Files/Practice-Parameters/qct.pdf?la=en
33. Çiçek Ö, Abdulkadir A, Lienkamp S, Brox T, Ronneberger O (2016) 3D U-Net_ Learning Dense Volumetric__Segmentation from Sparse Annotation. 9901. Springer, Cham. pp. 424–432
34. Huang G, Liu Z, van de Maaten L, Weinberger K (2017) Densely connected convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Publisher's Note