AI-Based Thoracolumbar Spine Bone Mineral Density Measurement System and Its Calibration Study Postprint

Authors: Xiong Xin, Li Yang, Shi Feng, Yang Lian, Duan Wei, Chen Bei, Li Yong, Zhao Linwei, Fu Quanshui, Fan Xiaoping, Yang Guoqing, Yang Guoqing

Date: 2025-03-20T00:00:00+00:00

Abstract

Background In recent years, the aging trend in China has gradually intensified, and the incidence of osteoporosis has also risen, becoming a major disease affecting the physical and mental health of the elderly. Moreover, the cost of diagnosing and treating osteoporosis is relatively high both domestically and internationally; therefore, early diagnosis of osteoporosis has become crucial for reducing patient suffering and treatment costs. Objective Based on conventional chest and abdominal CT plain scan images, to establish chest and abdominal bone density measurement models through deep neural networks and machine learning algorithms, and to calibrate chest bone density measurement results using the abdominal measurement model, thereby achieving automated bone density measurement and osteoporosis diagnosis. Methods A total of 702 patients who underwent both chest CT examination and QCT examination within a one-year interval between March 2022 and June 2023 at Suining Central Hospital in Sichuan Province were retrospectively collected as study subjects. Among them, 532 cases were randomly divided into a training set (426 cases, 80%) and a validation set (106 cases, 20%). The remaining 170 cases served as an internal test set for the model. This study used quantitative CT (QCT) diagnostic results as the reference standard and employed machine learning methods such as logistic regression, stochastic gradient descent, and random forest to construct osteoporosis classification models and bone density regression models for the chest and abdomen. The established models were internally tested, and classification performance was evaluated using metrics including sensitivity, specificity, accuracy, precision, and area under the receiver operating characteristic curve (AUC), while regression performance was assessed using mean absolute error, root mean square error, and coefficient of determination. Results The AUC values for the chest and abdominal osteoporosis classification models on the validation set were 0.948 and 0.968, respectively, and the mean absolute errors for

the bone density regression models were 10.534 and 9.449, respectively. In the internal test set, the AUC values for the classification models were 0.905 and 0.926, respectively, and the mean absolute errors for the regression models were 9.255 and 7.924, respectively. After calibration, the AUC and mean absolute error for the chest bone density measurement model on the validation set improved to 0.967 and 10.511, respectively. Conclusion The AI-based chest and lumbar bone density measurement results show high correlation and consistency with bone density measured by QCT and can effectively diagnose osteoporosis. The calibrated chest bone density measurement model further improved the model' s performance in diagnosis, providing substantial potential for the application and development of chest CT plain scans in opportunistic screening for osteoporosis.

Full Text

Research on the Measurement System and Calibration of Thoracolumbar Vertebral Density Based on Artificial Intelligence

Xin Xiong[12], Yang Li[4], Feng Shi[4], Lian Yang[12], Wei Duan[12], Bei Chen[12], Yong Li[2], Linwei Zhao[2], Quanshui Fu[2], Xiaoping Fan[2], Guoqing Yang[3]*

[1]College of Medical Imaging, North Sichuan Medical University, Nanchong 637000, China
[2]Radiology Department of Suining Central Hospital, Suining 629000, China
[3]Suining Traditional Chinese Medicine Hospital, Suining 629000, China
[4]Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, 200232, China

*Corresponding author: Guoqing Yang, Chief Physician; E-mail: snygq@163.com

---

Abstract

Background
As China' s aging population continues to grow, the incidence of osteoporosis has been steadily increasing, posing a significant health challenge for the elderly. Furthermore, the high costs associated with osteoporosis diagnosis and treatment both domestically and internationally underscore the critical importance of early detection as a key strategy to reduce patient suffering and healthcare expenditures.

Objective
This study aims to develop chest and abdominal bone mineral density (BMD) measurement models using conventional chest and abdominal CT scans through deep neural networks and machine learning algorithms. The abdominal BMD

model is subsequently employed to calibrate chest BMD measurements, enabling automated BMD assessment and osteoporosis diagnosis.

Methods
We retrospectively collected 702 patients from Suining Central Hospital in Sichuan Province who underwent both chest CT scans and quantitative CT (QCT) examinations between March 2022 and June 2023. Among these, 532 patients were randomly allocated to a training set (426 cases, 80%) and a validation set (106 cases, 20%). An additional 170 patients constituted the internal testing set. Using QCT diagnostic results as the reference standard, we constructed osteoporosis classification models and BMD regression models for both chest and abdomen using machine learning methods including logistic regression, stochastic gradient descent, and random forest. The established models underwent internal testing, with classification performance evaluated using sensitivity, specificity, accuracy, precision, and area under the receiver operating characteristic curve (AUC), while regression performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), and R-squared.

Results
In the validation set, the osteoporosis classification models achieved AUC values of 0.948 for the chest and 0.968 for the abdomen. The BMD regression models showed MAE values of 10.534 and 9.449, respectively. In the internal testing set, classification model AUC values were 0.905 and 0.926, while regression model MAE values were 9.255 and 7.924. After calibration, the chest BMD measurement model' s AUC and MAE in the validation set improved to 0.967 and 10.511, respectively.

Conclusion
AI-based chest and abdominal BMD measurements demonstrate high correlation and consistency with QCT measurements, effectively diagnosing osteoporosis. The calibrated chest BMD measurement model further enhances diagnostic performance, offering significant potential for the application of chest CT scans in opportunistic osteoporosis screening.

Keywords
Osteoporosis; Bone mineral density; CT plain scan; Deep learning; Machine learning

---

Osteoporosis is a chronic skeletal disease characterized by low bone mineral density (BMD) and deterioration of bone microstructure [1]. As China' s population aging intensifies, osteoporosis incidence continues to rise. With diagnosis and treatment costs being substantial both domestically and internationally—more than half of osteoporosis patients have never undergone screening [2]—early diagnosis has become crucial for reducing patient suffering and healthcare costs. Current early diagnosis relies primarily on BMD measurement using quantitative ultrasound (QUS), quantitative computed tomography (QCT),

and dual-energy X-ray absorptiometry (DXA) [3]. However, QUS suffers from low accuracy, QCT requires phantoms and complex post-processing, and DXA ignores skeletal morphology and structure while facing equipment shortages. Many scholars have proposed AI-based BMD measurement on CT plain scans to overcome these limitations, though most models undergo only simple validation, leaving diagnostic performance uncertain. According to the 2018 Chinese QCT Osteoporosis Diagnosis Guidelines [4], QCT measures true volumetric BMD, more sensitively reflecting osteoporosis changes and unaffected by spinal degeneration or vascular calcification. Therefore, this study uses QCT as the gold standard to develop separate chest and abdominal BMD measurement models based on CT plain scans through AI technology, with abdominal models calibrating chest measurements to enhance diagnostic performance and advance opportunistic screening using chest CT, thereby promoting AI development in osteoporosis diagnosis and providing new approaches for other disease models.

Methods

Study Population and Data Collection We retrospectively collected CT scans from 702 patients who underwent both chest CT and QCT examinations within approximately one year between March 2022 and June 2023 at Suining Central Hospital. Exclusion criteria included: (1) prior spinal surgery, (2) vertebral fractures, (3) localized vertebral density alterations from bone islands, cysts, or tumors, (4) severe spinal deformities such as scoliosis, and (5) poor CT image quality due to motion, respiratory, or metal artifacts. The study was approved by the Suining Central Hospital Ethics Committee (approval number: KYLLKS20230143).

Imaging Parameters Chest CT scanning parameters were 120 kV, 60 mA (automatic adjustment), 0.6 mm slice thickness, covering from lung apex to base. QCT parameters were 120 kV, 35 mA (automatic adjustment), 0.6 mm slice thickness. Lumbar CT data were processed through a CT workstation and transmitted to QCT-3000 bone densitometry software (Image Analysis, Inc.) for BMD calculation.

Image Acquisition and QCT Post-processing Following exclusion criteria, we retrieved eligible patients'chest CT plain scans and QCT raw images from the Picture Archiving and Communication System (PACS) and exported them in DICOM format, organized by patient name and scan region. All raw QCT images were reconstructed at the CT post-processing workstation. For each patient, we selected the central slices of L1-L3 vertebrae and measured the mean CT value of a standard density phantom using the largest possible ROI. Using an equally sized ROI, we then measured the mean CT value of vertebral trabecular bone, carefully avoiding bone cortex and basivertebral foramina. QCT-3000 Densitometry 2.0 software (Image Analysis, Inc.) calculated individual vertebral BMD and average BMD values. When fractured or pathological vertebrae

were present, adjacent vertebrae were measured instead. Quality control employed a calibration phantom composed of solid materials simulating vertebral components, with hydroxyapatite densities of 50 mg/cm^3, 100 mg/cm^3, and 200 mg/cm^3 in three standard vertebral bodies. Figure 1 [Figure 1: see original paper] illustrates the CT image acquisition and QCT post-processing workflow.

Clinical Data Collection We recorded patient examination numbers, sex, age, and QCT-measured BMD values. Diagnosis followed the 2018 Chinese QCT Osteoporosis Diagnosis Guidelines [4]: BMD $<$ 80 mg/cm^3 indicated osteoporosis, 80-120 mg/cm^3 indicated osteopenia, and $>$ 120 mg/cm^3 indicated normal bone mass.

Spine Segmentation and Feature Extraction De-identified CT plain scan sequences were processed using the uAI Research Portal (uRP) developed by Shanghai United Imaging Intelligence Co., Ltd. uRP is a medical image analysis software supporting visualization, automatic segmentation, and registration. Its built-in segmentation algorithm integrates a VB-Net pretrained model. VB-Net enhances the V-Net architecture with bottleneck structures, improving segmentation capabilities beyond traditional convolutional neural networks [5,6]. We applied uRP' s vertebral segmentation model to obtain automatic vertebral segmentation results, then performed 3 mm erosion to extract trabecular bone as the ROI for subsequent feature extraction. Figure 2 [Figure 2: see original paper] shows abdominal CT sagittal images and vertebral segmentation results.

BMD Measurement Model Construction The BMD measurement framework comprised both osteoporosis classification and BMD regression models. Using QCT measurements as the gold standard, we developed chest and abdominal models by extracting mean CT values (HU) from segmented vertebral ROIs and incorporating sex and age as clinical features. The modeling process involved: (1) calculating mean HU values for different vertebral ROIs, (2) using sex, age, and ROI HU means as features, and (3) inputting QCT-derived BMD classifications as labels into three algorithms: Bagging Decision Tree (BDT), Logistic Regression (LR), and Random Forest (RF) for classification; and BDT, Stochastic Gradient Descent (SGD), and RF for regression. All features were Z-score standardized before model input. Chest (T5-T10) and abdominal (T12-L2) models were constructed separately.

Model Calibration Since the gold standard BMD values were obtained from QCT measurements of L1-L3 or adjacent vertebrae, while the chest model (T5-T10) contained only thoracic vertebral features lacking lumbar information, we developed a thoracolumbar HU value calibration curve. We calculated mean HU values for each vertebra from T5 to L2, performed linear regression to obtain the average change coefficient k, and computed calibrated thoracic HU values (HUcal) using the intercept:

$$HUcal = k \times vertebran + Intercept$$

where k represents the HU change coefficient, vertebran is the vertebral level number (T5-L2: 1-10), Intercept is the intercept term, and HUcal is the calibrated vertebral HU value. This calibration method applied lumbar characteristic parameters to the chest model to improve its accuracy. Figure 3 [Figure 3: see original paper] presents the overall experimental workflow.

Statistical Analysis Model construction and statistical analysis were performed using Python 3.7.4 and R 4.1.2. Normally distributed continuous variables were expressed as mean ± standard deviation, with intergroup comparisons using independent samples t-tests. Non-normally distributed variables were expressed as median (P25, P75), with three-group comparisons using Kruskal-Wallis H tests. Categorical variables were expressed as frequencies and compared using ^{2} tests. Receiver operating characteristic (ROC) curves were plotted to calculate area under the curve (AUC), sensitivity, specificity, accuracy, precision, and F1 scores for classification models. Regression model performance was evaluated using mean absolute error (MAE), root mean squared error (RMSE), explained variance score (EVS), and R-squared. Pearson correlation analysis and Bland-Altman plots assessed agreement between model-predicted and QCT-measured BMD. All tests were two-tailed, with $P < 0.05$ considered statistically significant.

Results

Patient Demographics We collected CT scans from 702 patients. The 532 cases from March 2022 to February 2023 were randomly divided into training (n = 426, 80%) and validation (n = 106, 20%) sets, while 170 cases from March to June 2023 served as the internal testing set. Age and sex differed significantly among training, validation, and internal testing sets ($P < 0.05$), while BMD values showed no significant differences ($P > 0.05$) (Table 1).

Model Performance Osteoporosis Classification Models
The BDT model achieved optimal performance, with validation set AUC values of 0.948 for chest (T5-T10) and 0.968 for abdomen (T12-L2). After calibration (T5-T10-Cal), the chest classification model AUC improved to 0.967. Detailed performance metrics are presented in Table 2 , with ROC curves shown in Figure 4 [Figure 4: see original paper].

BMD Regression Models
The SGD model performed best, with validation set MAE values of 10.534 for chest and 9.449 for abdomen. After calibration (T5-T10-Cal), chest regression model MAE decreased to 10.511. Detailed metrics are provided in Table 3 , with correlation curves in Figure 5 [Figure 5: see original paper] and Bland-Altman

plots in Figure 6 [Figure 6: see original paper] (all differences statistically significant, $P < 0.05$).

Internal Testing Set Results

In the 170-case internal testing set, BDT classification models achieved AUC values of 0.905, 0.926, and 0.918 for T5-T10, T12-L2, and T5-T10-Cal, respectively (Table 4 ; ROC curves in Figure 7 [Figure 7: see original paper]). Regression models showed MAE values of 9.255 and 7.924 (Table 5 ; all $P < 0.05$). Pearson correlation analysis demonstrated strong correlations between all three regression models and thoracoabdominal measurements (Figure 8 [Figure 8: see original paper]).

Discussion

Few studies have established and calibrated chest and abdominal BMD measurement models using AI. Most existing research performs only simple validation of diagnostic models. PAN et al. [7] calculated mean CT values within cylindrical volumes of interest at each vertebral level (T12-L2), mapping them to QCT-measured BMD values via linear functions and demonstrating good correlation. JIN et al. [8] correlated thoracic vertebrae and first lumbar vertebra CT values with femoral neck, total hip, and L1-L4 t-scores, statistically proving positive correlations, particularly strong between CT values and L1/hip t-scores. LI et al. [9] measured mean lumbar CT HU values, compared them with BMD and DXA t-scores, and generated ROC curves validating model performance, showing good correlation between mean HU and BMD/T-scores. ROCK et al. [10] used AI algorithms to localize thoracic vertebrae from chest CT studies, automatically calculating mean HU values with kilovolt peak-dependent spectral correction, demonstrating moderate correlation with DEXA. FANG et al. [11] performed automatic vertebral segmentation using AI, employing convolutional neural networks for BMD calculation with QCT post-processing values as the standard, showing excellent correlation between automatic and manual L1-L4 segmentation. RÜHLING et al. [12] analyzed relationships between cervicothoracic vertebral vBMD values and lumbar (L1-L3) averages, calculating vertebral-specific osteoporosis thresholds via linear regression and demonstrating high correlations. While these studies achieved good diagnostic efficacy and advanced AI in healthcare, most focused solely on model establishment without calibration to improve accuracy. Our study addresses this gap by calibrating thoracolumbar BMD measurement models to enhance precision.

We developed AI-based chest and abdominal BMD measurement models with automated calibration, enabling fully automated BMD assessment and osteoporosis diagnosis. Through systematic data collection, classification, post-processing, and spine segmentation using the research platform, we applied AI to thoracoabdominal CT plain scans, constructing classification and regression models via decision trees, logistic regression, stochastic gradient descent, and random forests. Results demonstrated robust performance: validation set AUCs of 0.948 and 0.968 for chest and abdomen classification models, respectively; MAEs of

10.534 and 9.449 for regression models. After calibration, chest model AUC and MAE improved to 0.967 and 10.511. These AI-based measurements show high correlation and consistency with QCT, effectively diagnosing osteoporosis while offering substantial potential for opportunistic screening using chest CT.

Using QCT instead of conventional lumbar DXA as our reference standard provides more reliable model evaluation. Although DXA is the most common reference standard for osteoporosis diagnosis [13] and widely used to assess CT value efficacy [14-16], DXA-derived BMD values are not calculated from CT values and are susceptible to aortic calcification and spinal degeneration [17]. Our models demonstrate strong correlation and consistency with QCT-derived BMD, enabling high-precision vertebral BMD measurement.

Our thoracolumbar HU value calibration curve represents a novel approach. Since the chest model (T5-T10) contains only thoracic features without lumbar information, its accuracy was relatively lower. Calibration significantly improved chest model performance, a methodology not previously reported in the literature, providing new insights for future disease diagnosis model calibration studies.

Our machine learning and deep learning-based thoracolumbar BMD measurement models, enhanced through calibration, improve performance and demonstrate AI' s potential in disease diagnosis. The models accommodate chest, abdominal, and combined thoracoabdominal CT examinations, enabling fully automated vertebral BMD measurement. Compared with traditional QCT and DXA, our approach eliminates cumbersome workflows and post-processing—simply inputting CT images into integrated software yields results without additional X-ray exposure, facilitating early screening in asymptomatic patients and reducing economic burden. Model calibration further improves performance, offering new methodologies for other disease diagnosis models.

Limitations

First, this retrospective study may introduce selection bias; future prospective studies incorporating DXA are needed. Second, our single-center methodology requires multi-center validation with larger datasets to improve generalizability, robustness, and reproducibility. Third, models were developed using CT plain scans and cannot be directly applied to contrast-enhanced CT; future research should investigate BMD changes pre- and post-contrast and potentially explore MRI applications. Fourth, our calibration used simple linear fitting, which may require refinement considering additional factors. Fifth, age stratification was not performed during calibration; future studies should investigate age-specific thoracolumbar BMD measurement models and calibration approaches.

References

[1] CHENG X G, YUAN H S, CHENG J L, et al. Chinese expert consensus on the diagnosis of osteoporosis by imaging and bone mineral density[J]. Quant Imaging Med Surg, 2020, 10(10): 2066-2077. DOI:10.21037/qims-2020-16.

[2] SMITH A D. Screening of bone density at CT: an overlooked opportunity[J]. Radiology, 2019, 291(2): 368-369. DOI:10.1148/radiol.2019190434.
[3] SCHULTZ K, WOLF J M. Emerging technologies in osteoporosis diagnosis[J]. J Hand Surg Am, 2019, 44(3): 240-243. DOI:10.1016/j.jhsa.2018.07.006.
[4] Chinese QCT Osteoporosis Diagnosis Guidelines Working Group, CHENG X G, WANG L, et al. Chinese quantitative CT (QCT) osteoporosis diagnosis guidelines (2018)[J]. Chinese Journal of Osteoporosis, 2019, 25(6): 733-737. DOI:10.3969/j.issn.1006-7108.2019.06.001.
[5] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). October 25-28, 2016, Stanford, CA, USA. IEEE, 2016: 565-571. DOI:10.1109/3DV.2016.79.
[6] HAN M, YAO G, ZHANG W, et al. Segmentation of CT thoracic organs by multi-resolution VB-nets[D]. Proceedings of the 2019 challenge on segmentation of thoracic organs at risk in CT images.
[7] PAN Y L, SHI D J, WANG H Q, et al. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening[J]. Eur Radiol, 2020, 30(7): 4107-4116. DOI:10.1007/s00330-020-06679-y.
[8] YANG J R, LIAO M, WANG Y L, et al. Opportunistic osteoporosis screening using chest CT with artificial intelligence[J]. Osteoporos Int, 2022, 33(12): 2547-2561. DOI:10.1007/s00198-022-06491-y.
[9] LI Y L, WONG K H, LAW M W, et al. Opportunistic screening for osteoporosis in abdominal computed tomography for Chinese population[J]. Arch Osteoporos, 2018, 13(1): 76. DOI:10.1007/s11657-018-0492-y.
[10] SAVAGE R H, VAN ASSEN M, MARTIN S S, et al. Utilizing artificial intelligence to determine bone mineral density via chest computed tomography[J]. J Thorac Imaging, 2020, 35(Suppl 1): S35-S39. DOI:10.1097/RTI.0000000000000484.
[11] FANG Y J, LI W, CHEN X J, et al. Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks[J]. Eur Radiol, 2021, 31(4): 1831-1842. DOI:10.1007/s00330-020-07312-8.
[12] RÜHLING S, SCHAR A, SOLLMANN N, et al. Proposed diagnostic volumetric bone mineral density thresholds for osteoporosis and osteopenia at the cervicothoracic spine in correlation to the lumbar spine[J]. Eur Radiol, 2022, 32(9): 6207-6214. DOI:10.1007/s00330-022-08721-7.
[13] MICHAEL LEWIECKI E, GORDON C M, BAIM S, et al. International society for clinical densitometry 2007 adult and pediatric official positions[J]. Bone, 2008, 43(6): 1115-1121. DOI:10.1016/j.bone.2008.08.106.
[14] KIM Y W, KIM J H, YOON S H, et al. Vertebral bone attenuation on low-dose chest CT: quantitative volumetric analysis for bone fragility assessment[J]. Osteoporos Int, 2017, 28(1): 329-338. DOI:10.1007/s00198-016-3724-2.
[15] ROMME E A, MURCHISON J T, PHANG K F, et al. Bone attenuation on routine chest CT correlates with bone mineral density on DXA in patients with COPD[J]. J Bone Miner Res, 2012, 27(11): 2338-2343. DOI:10.1002/jbmr.1678.
[16] PICKHARDT P J, LEE L J, DEL RIO A M, et al. Simultaneous screening

for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard[J]. J Bone Miner Res, 2011, 26(9): 2194-2203. DOI:10.1002/jbmr.428.
[17] ENGELKE K, LIBANATI C, LIU Y, et al. 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)[J]. Bone, 2009, 45(1): 110-118. DOI:10.1016/j.bone.2009.03.669.

Received: July 10, 2024; Revised: January 25, 2025
Edited by: Shiyue Wang

Note: Figure translations are in progress. See original paper for figures.