Feasibility of Bone Mineral Density and Bone Microarchitecture Assessment Using Deep Learning With a Convolutional Neural Network

Kazuki Yoshida, MD, PhD, Yuki Tanabe, MD, PhD,* Hikaru Nishiyama, MS,* Takuya Matsuda, MD, PhD,†
Hidetaka Toritani, MS,§ Takuya Kitamura, MD,* Shinichiro Sakai, MD,‡ Kunihiko Watamori, MD, PhD,‡
Masaki Takao, MD, PhD,‡ Eizen Kimura, BM, PhD,† and Teruhito Kido, MD, PhD**

Objectives: We evaluated the feasibility of using deep learning with a convolutional neural network for predicting bone mineral density (BMD) and bone microarchitecture from conventional computed tomography (CT) images acquired by multivendor scanners.

Methods: We enrolled 402 patients who underwent noncontrast CT examinations, including L1–L4 vertebrae, and dual-energy x-ray absorptiometry (DXA) examination. Among these, 280 patients (3360 sagittal vertebral images), 70 patients (280 sagittal vertebral images), and 52 patients (208 sagittal vertebral images) were assigned to the training data set for deep learning model development, the validation, and the test data set, respectively. Bone mineral density and the trabecular bone score (TBS), an index of bone microarchitecture, were assessed by DXA. BMD_{DL} and TBS_{DL} were predicted by deep learning with a convolutional neural network (ResNet50). Pearson correlation tests assessed the correlation between BMD_{DL} and BMD, and TBS_{DL} and TBS. The diagnostic performance of BMD_{DL} for osteopenia/osteoporosis and that of TBS_{DL} for bone microarchitecture impairment were evaluated using receiver operating characteristic curve analysis.

Results: BMD_{DL} and BMD correlated strongly ($r = 0.81$, $P < 0.01$), whereas TBS_{DL} and TBS correlated moderately ($r = 0.54$, $P < 0.01$). The sensitivity and specificity of BMD_{DL} for identifying osteopenia or osteoporosis were 93% and 90%, and 100% and 94%, respectively. The sensitivity and specificity of TBS_{DL} for identifying patients with bone microarchitecture impairment were 73% for all values.

Conclusions: The BMD_{DL} and TBS_{DL} derived from conventional CT images could identify patients who should undergo DXA, which could be a gatekeeper tool for detecting latent osteoporosis/osteopenia or bone microarchitecture impairment.

O steoporosis is a common and treatable skeletal disease caused by decreased bone mineral density (BMD) and bone microarchitecture impairment.[1] Osteoporosis increases the probability of fragility fractures and the mortality rate. Therefore, an early diagnosis of osteoporosis is important to prevent fragility fractures.[2] Current guidelines recommend BMD assessment by using dual-energy x-ray absorptiometry (DXA) to diagnose osteoporosis, as BMD is associated with approximately 70% bone strength.[3,4] However, more than half of fragility fractures occur despite normal BMD.[3–5] Bone microarchitecture is another important factor in fragility fractures. The trabecular bone score (TBS) indicates bone microarchitecture and is derived from assessing DXA images and performing texture analysis.[3,6,7] A lower TBS increases the probability of fragility fractures, independent of BMD.[8,9] Recent guidelines state that the TBS can be applied to the risk assessment of fragility fractures.[3]

There are many undetected patients with osteoporosis in the real world, as DXA is not often performed for asymptomatic patients.[1] Computed tomography (CT) has been widely used in clinical practice, and it might be applied as a gatekeeper tool for detecting latent osteoporosis. A previous report showed that CT attenuation (as measured in Hounsfield units [HUs]) correlated moderately with BMD.[10] Recently, deep learning has been applied to assess osteoporosis.[11,12] However, the usefulness of deep learning for bone microarchitecture assessment is not clear. We, therefore, aimed to investigate the feasibility of predicting BMD and TBS from abdominal CT images by using deep learning with a convolutional neural network (CNN).

MATERIALS AND METHODS

Study Population

This study was a retrospective cross-sectional study approved by the local institutional review board (approval number: 2105008). The need for informed consent was waived. We retrospectively identified the study patients between July 2016 and March 2021 (age ≥20 years). We included the patients who underwent CT examination (120 kVp), including the lumbar spine (L1–L4), and lumbar DXA examination within 1 year. We determined the acceptable time interval between DXA and CT in inclusion criteria according to the previous report.[1,12] We excluded patients with a lumbar compression fracture, severe scoliosis, severe spondylosis, previous lumbar spine surgery, and enhanced CT to avoid the change in the CT attenuation of the lumbar vertebra.[12] Four hundred two patients who matched the criteria were enrolled in this study. Among

TABLE 1. Patient Characteristics

	Training + Validation, n = 350	Test, n = 52	*P*
Age, y	63 (49–71)	65 (51–74)	0.50
Men	80 (23%)	11 (21%)	0.78
Body mass index, kg/m^2	21.9 (19.6–24.9)	21.9 (19.5–25.9)	0.64
Time interval between the DXA and CT, d	49 (9–121)	87 (14–185)	0.12
CT vendor			0.32
Canon Medical Systems	155 (44%)	18 (35%)	
Philips Healthcare	81 (23%)	12 (23%)	
Siemens Healthineers	114 (33%)	22 (42%)	

Continuous data are presented as the median (25th–75th percentile) and assessed by Wilcoxon signed rank test.

Number (%) of subjects are assessed by the χ^2 test (in the case of 3 groups, using the Bonferroni correction).

*Statistical significance was determined at $P < 0.05$ between the training + validation data set and test data set.

these patients, case numbers (1–402) were randomly assigned to the consecutive patient data set. We divided into 1–350 as the training and validation data set (350 patients) and 351–402 as the test data set (52 patients, 208 CT images). In each fold of 5-fold cross-validation, 280 patients (3360 CT images) were assigned to the training data set for model development, and 70 patients (280 CT images) were assigned to the validation data set. There was no overlap between the training data set and the test data set. Patient characteristics and information about the patient's indications for imaging are shown in Table 1 and 2.

CT Acquisition

Computed tomography images were acquired by a 320-row multidetector row volume CT scanner (Aquilion ONE; Canon Medical Systems, Otawara, Japan), a third-generation dual-source CT scanner (SOMATOM Force; Siemens Healthineers, Erlangen, Germany), and a 256-slice multidetector row CT scanner (Brilliance iCT; Philips Healthcare, Best, the Netherlands). The scan parameters and reconstruction technique from each CT scanner are shown in Table 3.

Input Image Data

We first performed segmentation of the L1–L4 vertebra semiautomatically using a bone extraction application from a dedicated workstation (Synapse Vincent; Fuji Medical Systems, Minato, Japan). For the training data sets, we manually cropped 3 sagittal cross-sectional images at and close (5 mm gap) to the midvertebra (5-mm thickness, field of view [FOV] 300 × 300 mm) from each L1–L4 lumbar vertebra (Fig. 1). For the test and validation data sets, we manually cropped 1 sagittal cross-sectional image (5-mm thickness, FOV 300 × 300 mm) at midvertebra from each L1–L4 lumbar vertebra.[12] Finally, we obtained CT images from the workstation in the Digital Images and Communications in Medicine (DICOM) format. The DICOM format images were then converted to JPEG images and resized to 256 × 256 pixels using the Python 3.5.4 programming language (library; os, pydicom, cv2, pandas, shutil).

BMD and TBS Prediction Using Deep Learning With a CNN

We used deep learning with a CNN (ResNet50) to predict BMD and the TBS. The preprocessed image data were inputted into the deep learning algorithm and augmented by image augmentation layer. The schema of the deep learning algorithm (ResNet50) is illustrated in Figure 2. The hyperparameters were as follows: the number of epochs, 500; optimizer, Adam[12]; and minibatch size,

20 determined by GPU performance. We used to save the best mechanism in the Neural Network Console and adopted the model with the lowest validation loss. The image augmentation and Huber loss function layers are active only in the training process. In the training and validation process, 5-fold cross-validation was performed. First, we excluded the test data from all data sets. The remaining data were divided randomly into 5 groups. The untrained model was trained on 4 groups and validated one. Then, the procedure was repeated 5 times to complete validation on all 5 groups. Finally, we assessed the trained model performance in the test data set. For vertebra-based analysis, BMD acquired by deep learning with the CNN (BMD$_{DL}$) and the TBS acquired by deep learning with each vertebra's CNN (TBS$_{DL}$) from test data sets were used. In addition, for patient-based analysis, the mean of the L1–L4 lumbar vertebra was used as patient-based BMD$_{DL}$ and TBS$_{DL}$.

The software to build the models was neural network console version 1.9.7587.58782 (Sony Network Communications Inc, Shinagawa, Japan). The whole process was run on a Core i7-6800K central processing unit (Intel), 3.4 GHz with a GeForce GTX 1080Ti graphics processing unit (NVIDIA), and 32 GB of random-access memory.

CT Attenuation Measurements in the Test Data Set

For vertebra-based analysis, a diagnostic radiologist with 8 years of experience independently placed a circular region of interest (100–150 mm^2) on the center of the L1–L4 lumbar vertebra and assessed the CT attenuation (in HU). For patient-based analysis,

TABLE 2. Indications for CT Examination

	No. patients (%)
Orthopedic disease	111 (28%)
Autoimmune disease	110 (27%)
Gastrointestinal disease	78 (19%)
Metabolic and endocrine disease	27 (7%)
Breast disease	24 (6%)
Neuromuscular disease	14 (3%)
Lung disease	12 (3%)
Hematology and infection disease	8 (2%)
Nephrology and urology disease	6 (1%)
Gynecology disease	6 (1%)
Others	6 (1%)

Data are presented as number (%) of subjects.

TABLE 3. Scan and Reconstruction Parameters of CT Examinations

CT Scanner Vendor	Aquilion ONE (n = 173) Canon Medical Systems	SOMATOM Force (n = 136) Siemens Healthineers	Brilliance iCT (n = 93) Philips Healthcare
Tube voltage, kVp	120	120	120
Reconstruction algorithm	AIDR3D strong	ADMIRE level 2 or 3	iDose 4 level5
Kernel for reconstruction	FC03, FC18	Br40	B, C
Slice thickness, mm	0.5–1.0	0.6–1.0	0.67–1.0
Field of view, mm	349.2–497.7	350–500	350–500

the mean of L1–L4 lumbar vertebral CT attenuation was used as the patient-based CT attenuation of the lumbar vertebra.

Dual-Energy X-ray Absorptiometry

Dual-energy x-ray absorptiometry images were acquired with a Lunar Prodigy (GE Healthcare, Little Chalfont, United Kingdom). Bone mineral density and the % young adult mean (%YAM) calculated for the lumbar vertebra (L1–L4) were recorded from the reporting system. In addition, the TBS of the lumbar vertebra (L1–L4) was calculated by TBS iNsight software (Version 3.0.3.0; Medimaps, Bordeaux, France), by a diagnostic radiologist with 12 years of experience.[9] The mean of the L1–L4 lumbar vertebra was used as the patient-based BMD and TBS. Patients were diagnosed with osteopenia (%YAM <80%) or osteoporosis (%YAM <70%), respectively.[3,4] Patients were also diagnosed with bone microarchitecture impairment when the patient-based TBS was ≤1.31.[13]

Statistical Analysis

We used the Shapiro-Wilk test to evaluate the normality of data distributions. Continuous data are expressed as the mean (SD) or median (25th–75th percentile), and assessed by Student *t* test or Wilcoxon signed rank test, as appropriate. The categorical variables are expressed as percentages and assessed by the χ^2 test (in the case of 3 groups, using the Bonferroni correction).

We used the Pearson correlation test to assess the correlations between the BMD_{DL} and BMD, TBS_{DL} and TBS, CT attenuation and BMD, or CT attenuation and TBS in each vertebra. The differences between the BMD_{DL} and BMD, TBS_{DL} and TBS, patient-based BMD_{DL} and patient-based BMD, or patient-based TBS_{DL} and patient-based TBS were compared by the Student *t* test. We also performed subgroup analysis (eg, CT vendor, sex, age, body mass index [BMI], and time interval between DXA and CT). In the subgroup analysis, we divided the test patients into 3 groups according to the CT vendor or 2 groups according to sex, median age, median BMI, and the median time interval between DXA and CT interval, respectively. In addition, multivariable linear regression analysis was performed to evaluate the factors affecting on the BMD or TBS prediction with deep learning among the subgroup.

We used receiver operating characteristic curve analyses to evaluate the diagnostic performance of patient-based BMD_{DL} for osteopenia/osteoporosis and patient-based TBS_{DL} for bone microarchitecture impairment.[14] We calculated the areas under the curve (AUCs). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), false-predictive rate (FPR), false-negative predictive rate (FNR), and accuracy were calculated at a patient level.
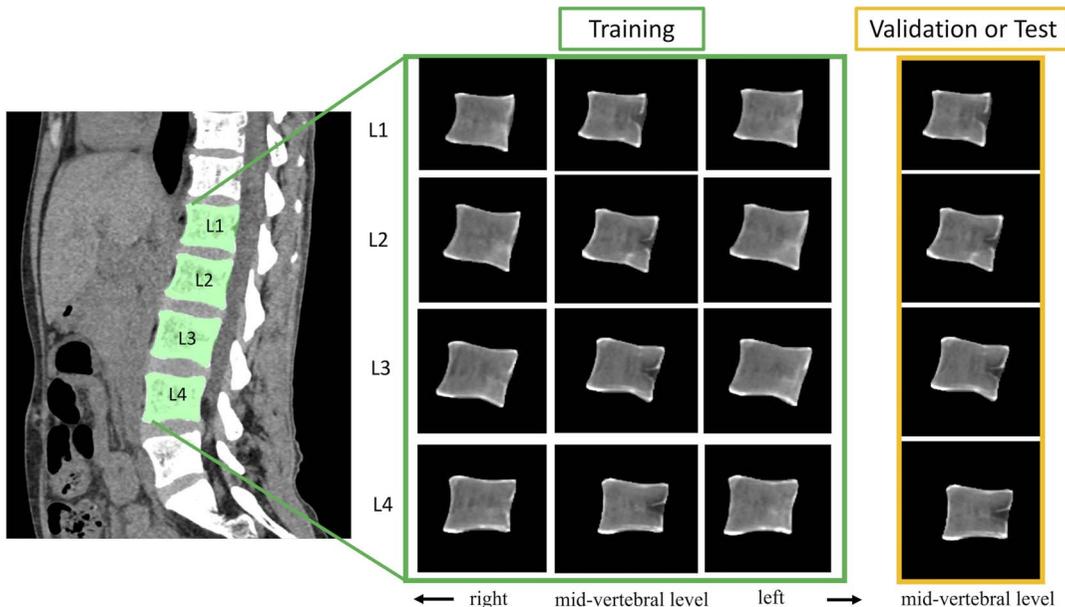
FIGURE 1. Preprocessing of input images. We cropped sagittal cross-sectional images of each lumbar vertebra (5-mm thickness, 512 × 512 pixels, FOV 300 × 300 mm). We cropped 3 sagittal cross-sectional images at and close to the midvertebrae from each L1–L4 lumbar vertebra for the training data sets and 1 sagittal cross-sectional image at midvertebrae from each L1–L4 lumbar vertebra for the validation and test data sets. Figure 1 can be viewed in color online at www.jcat.org.

FIGURE 2. Schema of the deep leaning with CNN (ResNet50). We present the data shape (number, x-size, and y-size of image/feature map) below the Input data, IA, Conv, and MP layer. We also exhibit the number of calculated value below the Affine and Huber Loss layer. CNN, convolutional neural network; IA, image augmentation; Conv, convolution; MP, max pooling.

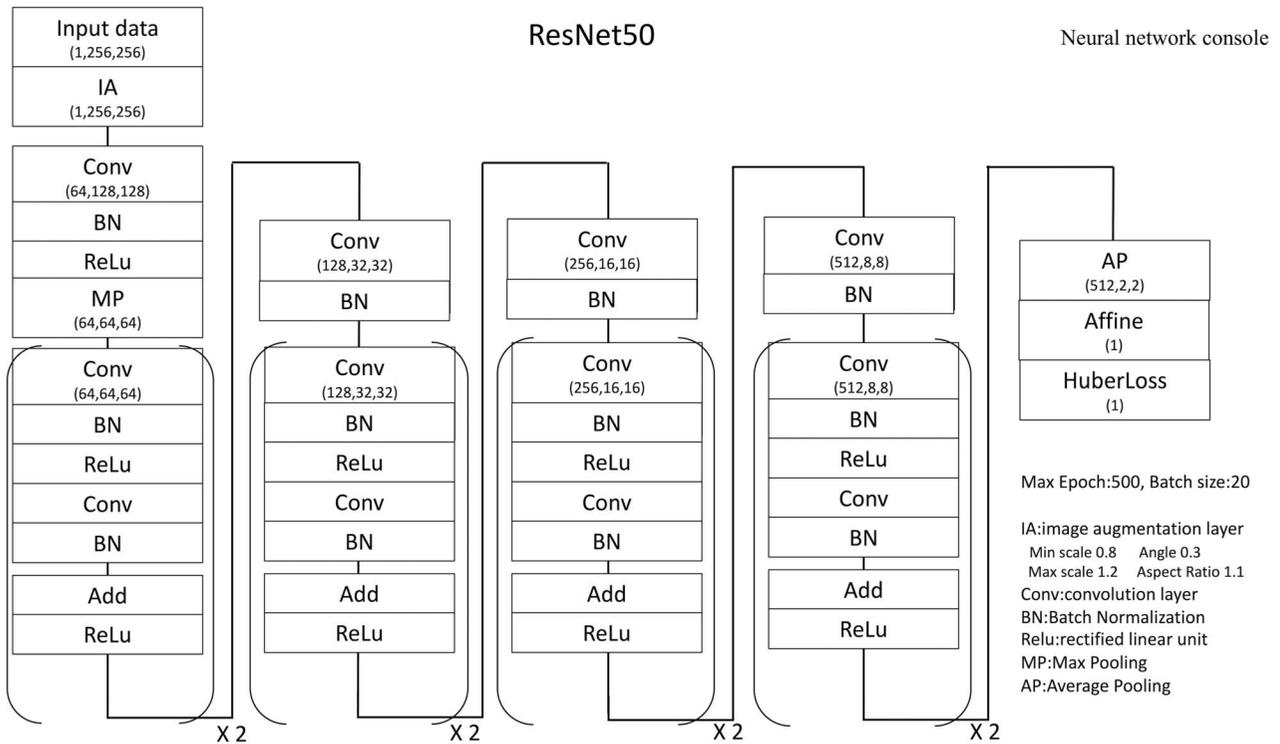We calculated the sample size for the test data set based on BMD because there were few previous reports associated with TBS predicted by deep learning. We assumed the standard deviation of BMD was 0.11 g/cm^2,[15] the statistical power was 0.8, and a 2-sided significance level was 0.05. The estimated BMD difference between DXA and deep learning was determined based on previous research.[16] Thirty-four patients and more were needed to detect more than 0.11 g/cm^2 absolute difference for BMD measurements. We also calculated the statistical power of TBS prediction based on our results and reported SD = 0.10 of TBS from a previous study.[9]

In all tests, statistical significance was set at $P < 0.05$. Statistical analyses were performed with JMP14 (SAS Institute Inc, Cary, NC) and SPSS ver. 26 software (IBM SPSS, Chicago, IL).

RESULTS

DXA in the Test Data Set

The BMD values and TBS values were 1.02 (0.22) g/cm^2 and 1.36 (0.13) per vertebra, respectively. The patient-based BMD and TBS were 1.03 (0.20) g/cm^2 and 1.36 (0.11) per patient, respectively. Among the 52 patients, 17 and 4 patients were diagnosed with osteopenia and osteoporosis, respectively. In addition, 15 of the 52 patients were diagnosed as having bone microarchitecture impairment.

Correlation and Comparison Between BMD$_{DL}$ and BMD, and Between TBS$_{DL}$ and TBS in the Per-Vertebra Analysis in the Test Data Set

The BMD$_{DL}$, TBS$_{DL}$, and CT attenuation were 1.07 (0.19) g/cm^2, 1.36 (0.09), and 132 (58) HU, respectively, in the test data set.

There was a significantly strong correlation between the BMD$_{DL}$ and BMD ($r = 0.81$, $P < 0.01$) (Fig. 3A). There was a significant moderate correlation between CT attenuation and BMD ($r = 0.60$, $P < 0.01$) (Fig. 3B). In addition, there was a moderate correlation between the TBS$_{DL}$ and TBS ($r = 0.54$, $P < 0.01$) (Fig. 3C). The correlation between CT attenuation and the TBS was weak ($r = 0.25$, $P < 0.01$) (Fig. 3D). BMD$_{DL}$ was significantly higher than BMD ($P = 0.03$), and there was no significant difference between TBS$_{DL}$ and TBS ($P = 0.81$). The statistical power of TBS at a vertebral level was 0.99.

In the subgroup analysis associated with CT vendor, sex, age, BMI, and time between DXA and CT, there was a significantly moderate-strong correlation between the BMD$_{DL}$ and BMD in all subgroups: vendor A ($r = 0.83$, $P < 0.001$), vendor B ($r = 0.84$, $P < 0.001$), and vendor C ($r = 0.65$, $P < 0.01$); males ($r = 0.91$, $P < 0.001$) and females ($r = 0.78$, $P < 0.001$); young ($r = 0.72$, $P < 0.001$) and seniors ($r = 0.83$, $P < 0.001$); low BMI ($r = 0.80$, $P < 0.001$) and high BMI ($r = 0.81$, $P < 0.001$); short time interval ($r = 0.81$, $P < 0.001$) and long time interval ($r = 0.80$, $P < 0.001$), respectively. In addition, there was a significant moderate correlation between the TBS$_{DL}$ and TBS in all subgroups: vendor A ($r = 0.46$, $P < 0.001$), vendor B ($r = 0.57$, $P < 0.001$), and vendor C ($r = 0.53$, $P < 0.01$); males ($r = 0.62$, $P < 0.001$) and females ($r = 0.52$, $P < 0.001$); young ($r = 0.54$, $P < 0.001$) and seniors ($r = 0.47$, $P < 0.001$); low BMI ($r = 0.68$, $P < 0.001$) and high BMI ($r = 0.45$, $P < 0.001$); short time interval ($r = 0.53$, $P < 0.001$) and long time interval ($r = 0.53$, $P < 0.001$), respectively.

The BMD$_{DL}$, BMD, TBS$_{DL}$, and TBS in the subgroup analysis were shown in Table 4. There were significant differences between the BMD$_{DL}$ and BMD in vendor C, female, low BMI, and short time interval. There were no significant differences between the TBS$_{DL}$ and TBS in all subgroups. In the multivariable linear

regression analysis, CT vendor and BMI were significant sub-group factors affecting the BMD prediction with deep learning, whereas there was no significant subgroup factor affecting the TBS prediction with deep learning (Table 5).

Diagnostic Performance of Patient-Based BMD$_{DL}$ and TBS$_{DL}$ Per-Patient Analysis in the Test Data Set

The patient-based BMD$_{DL}$, patient-based TBS$_{DL,}$ and patient-based CT attenuation were 1.07 (0.17) g/cm^2, 1.36 (0.05), and 132 (56) HU, respectively, in the test data set. All of them showed normal distribution. The AUCs of patient-based BMD$_{DL}$ for identifying osteopenia and osteoporosis were 0.921 (95% confidence interval [CI], 0.793–0.973) and 0.969 (95% CI, 0.872–0.993), respectively. The sensitivity, specificity, PPV, NPV, FPR, NPR, and accuracy of patient-based BMD$_{DL}$ for identifying osteopenia and osteoporosis were 93%, 90%, 77%, 97%, 11%, 7%, and 90%, and 100%, 94%, 57%, 100%, 6%, 0%, and 94%, respectively. In addition, the AUC of patient-based TBS$_{DL}$ for identifying patients with bone microarchitecture impairment was 0.768 (95% CI, 0.585–0.886). The sensitivity, specificity, PPV, NPV,

FPR, NPR, and accuracy of patient-based TBS$_{DL}$ for identifying patients with bone microarchitecture impairment were 73%, 73%, 52%, 87%, 27%, 27%, and 73%, respectively. There were no significant differences between patient-based BMD$_{DL}$ and patient-based BMD ($P = 0.26$), or between patient-based TBS$_{DL}$ and patient-based TBS ($P = 0.88$). The statistical power of TBS at a patient level was 0.98.

DISCUSSION

In this study, we found a strong correlation between BMD$_{DL}$ and BMD, and a moderate correlation between TBS$_{DL}$ and TBS. We showed that deep learning with CNN could identify patients with osteopenia/osteoporosis or bone microarchitecture impairment.

The present study showed that deep learning with CNN allowed accurate prediction of BMD from abdominal CT images. Computed tomography attenuation also presented a significant correlation with BMD, but the correlation with BMD was weaker than that with BMD$_{DL}$. Several previous studies also reported that deep learning with a CNN could predict BMD from noncontrast LI–L4 CT images, but the deep learning algorithms in those studies

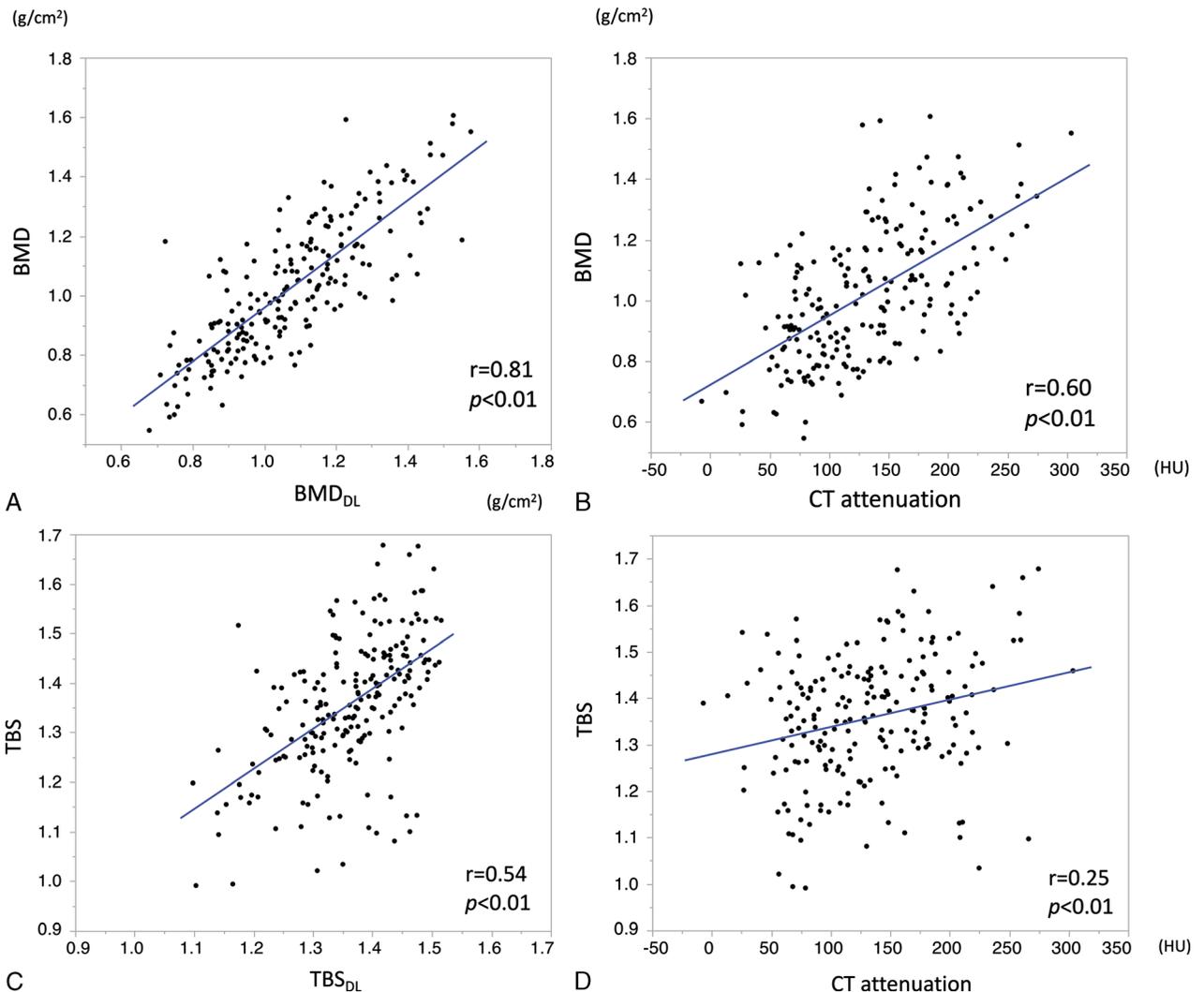**FIGURE 3.** Correlations between BMD$_{DL}$ and BMD (A), CT attenuation and BMD (B), TBS$_{DL}$ and TBS (C), and CT attenuation and TBS (D) in test data sets. DL indicates deep learning–predicted. Figure 3 can be viewed in color online at www.jcat.org.

TABLE 4. Subgroup Results of BMD and TBS Prediction Per-Vertebra Analysis

	BMD	BMD$_{DL}$	*P*
CT vendor			
Vendor A	1.05 (0.19)	1.03 (0.24)	0.47
Vendor B	1.21 (0.20)	1.19 (0.20)	0.60
Vendor C	1.01 (0.16)	0.93 (0.15)	<0.01*
Sex			
Male	1.10 (0.21)	1.06 (0.22)	0.36
Female	1.06 (0.19)	1.02 (0.22)	0.04*
Age			
Young	1.13 (0.18)	1.09 (0.18)	0.13
Senior	1.00 (0.20)	0.95 (0.23)	0.07
Body mass index			
Low	1.05 (0.21)	0.97 (0.23)	0.01*
High	1.09 (0.18)	1.08 (0.19)	0.56
Time interval			
Short	1.09 (0.19)	1.03 (0.21)	0.04*
Long	1.05 (0.20)	1.02 (0.22)	0.26
	TBS	**TBS$_{DL}$**	**P**
CT vendor			
Vendor A	1.35 (0.09)	1.36 (0.14)	0.81
Vendor B	1.40 (0.09)	1.43 (0.13)	0.21
Vendor C	1.35 (0.09)	1.32 (0.12)	0.10
Sex			
Male	1.36 (0.09)	1.35 (0.12)	0.77
Female	1.36 (0.09)	1.36 (0.14)	0.91
Age			
Young	1.38 (0.09)	1.39 (0.13)	0.45
Senior	1.34 (0.09)	1.32 (0.13)	0.25
Body mass index			
Low	1.36 (0.09)	1.35 (0.11)	0.58
High	1.36 (0.09)	1.37 (0.15)	0.89
Time interval			
Short	1.37 (0.09)	1.36 (0.13)	0.38
Long	1.35 (0.09)	1.36 (0.14)	0.59

Continuous data are expressed as the mean (SD) and assessed by Student *t* test.

*Statistical significance was determined at *P* < 0.05 between the BMD and BMD$_{DL}$.

BMD$_{DL}$, BMD acquired by deep learning with the CNN; TBS$_{DL}$, TBS acquired by deep learning with the CNN.

were trained using a single-vendor CT scanner data set.[11,12] Recently, it was recommended that multivendor images should be used in each step of the deep learning algorithm development to prevent overfitting. Vendor-specific deep learning algorithms are much less useful for clinical situations than multivendor adaptable deep learning algorithms.[17] In the present study, conventional CT images acquired by multivendor scanners were used in all phases of the deep learning algorithm evaluation. Our results showed the robustness of BMD prediction using deep learning with a CNN even in a data set acquired with multivendor CT scanners. In the bone microarchitecture assessment, the TBS assessed clinically by DXA was used for identifying bone microarchitecture impairment.[3,6,7] Deep learning with a CNN allowed prediction of the TBS from abdominal CT images, whereas CT attenuation

exhibited a weak correlation with the TBS in the present study. Previous studies also showed that bone microarchitecture could be evaluated using CT images and structural/textural analysis.[6,18,19] We speculate that deep learning with CNN might predict bone microarchitecture concerning structural and textural characteristics, which was difficult to be evaluated by CT attenuation. However, the TBS$_{DL}$ did not match fully with the TBS assessed by DXA. We speculated that this was because the spatial resolution of CT images was not sufficient for the accurate prediction of the TBS.[7]

The present study showed that the BMD$_{DL}$ tended to be slightly higher than BMD. In the subgroup analysis, there were significant differences between the BMD$_{DL}$ and BMD in CT vendor, sex, BMI, and time interval. Among them, CT vendor and BMI significantly affected BMD prediction according to the multivariable linear regression analysis. In regards to the CT vendor, we speculated that the differences in effective energies and image noise due to the scan parameters and iterative reconstruction methods among CT vendors may have affected the prediction of BMD with deep learning.[20–22] A larger number of training data set was needed to further improve the accuracy of BMD prediction with deep learning in multivendor CT data set.[23] In regards to the BMI, we speculated that the differences in soft tissue information between CT and DXA. In the present study, CT contained only bone information from the cropped lumbar vertebra, whereas DXA contained information on bone, soft tissue, aorta, and other organs.[11,24] In addition, the dose adjustment mechanism based on body size may have affected the prediction of BMD with deep learning. Computed tomography had an autoexposure control mechanism that adjusted dose modulation based on body size,[25] whereas DXA did not have a dose modulation mechanism. Although these factors might affect BMD prediction, BMD$_{DL}$ presented high accuracy in identifying osteopenia/osteoporosis in this study, as shown in previous reports,[11,12] and it could be feasible to use BMD$_{DL}$ in clinical practice.

In addition to BMD, bone microarchitecture is also very important, as more than half of fragility fractures occur despite BMD lying within the normal range.[3,5,8,9,13] The TBS is an indirect measurement of bone microarchitecture, which is calculated by texture analysis from DXA images. Previous reports showed that the TBS correlated significantly with the 3D bone microarchitecture parameters in human cadavers and that the TBS could independently predict fragility fracture risk from BMD or Fracture Risk Assessment Tool.[8,9,13] The present study showed that no significant difference

TABLE 5. Multivariable Linear Regression Analysis for Evaluating Factors Associated With BMD or TBS Prediction With Deep Learning

Subgroup factors	*P*
BMD	
CT vendor	0.03*
Sex	0.50
Age	0.36
Body mass index	<0.001*
Time interval	0.15
TBS	
CT vendor	0.15
Sex	0.99
Age	0.07
Body mass index	0.53
Time interval	0.13

*Statistical significance was determined at *P* < 0.05.

exists between the TBS$_{DL}$ and TBS. According to the subgroup analysis, CT vendor, sex, age, BMI, and the time interval between DXA and CT might have no significant effect on the TBS prediction with deep learning. Although we could not calculate the adequate sample size associated with TBS prediction with deep learning due to the limits of the previous reports, the statistical power was high at both vertebral and patient levels. Moreover, the TBS$_{DL}$ derived from conventional CT images could identify bone microarchitecture impairment. The predictability of bone microarchitecture impairment by deep learning with a CNN will be improved by further developments in CT technology (eg, spatial resolution) in the future.

There were several limitations in the present study. First, this was a single-center retrospective cross-sectional study, and the number of the data set was relatively small. Recently, external validation is preferable to apply deep learning models in the real world, when possible.[22] Unfortunately, this study did not include external validation. Further prospective multicenter studies with large data sets and external validation are needed. Second, we are concerned that the manual part in the segmentation process might impair the reproducibility of the cropped vertebral images and the accuracy of the BMD and TBS prediction with deep learning in the present study. Recently, U-net could automatically segment and label vertebrae,[11] which would help us to improve our deep learning algorithm. Third, the effect of contrast media is not known. The present study did not include enhanced CT, to avoid the change in the CT attenuation of the lumbar vertebra. Fourth, the differences between other deep learning algorithms have not been investigated. AlexNet, GoogLeNet, and ResNet are often used as CNN models for diagnostic imaging analysis. Herein, we used ResNet50 because it exhibited the best diagnostic performance in a certain task.[26,27] Fifth, the time interval between the DXA and CT did not appear to match between the test data set of 87 (14–185) days and the training and validation data set of 49 (9–121) days. Although this difference was not statistically significant, we could not rule out the possibility of affecting the performance of the prediction of BMD and TBS with deep learning in this study. Finally, we did not evaluate the performance of deep learning with a CNN in the prediction of fragility fractures. The feasibility of BMD$_{DL}$ and TBS$_{DL}$ derived from conventional CT images for predicting fragility fractures should be evaluated in further studies.

In conclusion, deep learning with a CNN could predict both BMD and the TBS using conventional CT images acquired by multivendor scanners. The BMD$_{DL}$ and TBS$_{DL}$ derived from conventional CT images could identify patients who should undergo DXA, which could be a gatekeeper tool for detecting latent osteoporosis/osteopenia or bone microarchitecture impairment.

REFERENCES

1. Compston JE, McClung MR, Leslie WD. Osteoporosis. *Lancet*. 2019;393: 364–376.

2. Cosman F, Miller PD, Williams GC, et al. Eighteen months of treatment with subcutaneous abaloparatide followed by 6 months of treatment with alendronate in postmenopausal women with osteoporosis: results of the ACTIVExtend trial. *Mayo Clin Proc*. 2017;92:200–210.

3. Kanis JA, Cooper C, Rizzoli R, et al. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int*. 2019;30:3–44.

4. Orimo H, Nakamura T, Hosoi T, et al. Japanese 2011 guidelines for prevention and treatment of osteoporosis–executive summary. *Arch Osteoporos*. 2012;7:3–20.

5. Siris ES, Chen YT, Abbott TA, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med*. 2004; 164:1108–1112.

6. Link TM. Osteoporosis imaging: state of the art and advanced imaging. *Radiology*. 2012;263:3–17.

7. Hans D, Barthe N, Boutroy S, et al. Correlations between trabecular bone score, measured using anteroposterior dual-energy x-ray absorptiometry acquisition, and 3-dimensional parameters of bone microarchitecture: an experimental study on human cadaver vertebrae. *J Clin Densitom*. 2011;14: 302–312.

8. Hans D, Goertzen AL, Krieg MA, et al. Bone microarchitecture assessed by TBS predicts osteoporotic fractures independent of bone density: the Manitoba study. *J Bone Miner Res*. 2011;26:2762–2769.

9. Iki M, Tamaki J, Kadowaki E, et al. Trabecular bone score (TBS) predicts vertebral fractures in Japanese women over 10 years independently of bone density and prevalent vertebral deformity: the Japanese population-based osteoporosis (JPOS) cohort study. *J Bone Miner Res*. 2014;29:399–407.

10. Gausden EB, Nwachukwu BU, Schreiber JJ, et al. Opportunistic use of CT imaging for osteoporosis screening and bone density assessment: a qualitative systematic review. *J Bone Joint Surg Am*. 2017;99: 1580–1590.

11. Pan Y, Shi D, Wang H, et al. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. *Eur Radiol*. 2020;30:4107–4116.

12. Yasaka K, Akai H, Kunimatsu A, et al. Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. *Eur Radiol*. 2020;30: 3549–3557.

13. McCloskey EV, Odén A, Harvey NC, et al. A meta-analysis of trabecular bone score in fracture risk prediction and its relationship to FRAX. *J Bone Miner Res*. 2016;31:940–948.

14. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163:670–675.

15. Lee KS, Bae SH, Lee SH, et al. New reference data on bone mineral density and the prevalence of osteoporosis in Korean adults aged 50 years or older: the Korea National Health and nutrition examination survey 2008–2010. *J Korean Med Sci*. 2014;29:1514–1522.

16. Uemura K, Otake Y, Takao M, et al. Development of an open-source measurement system to assess the areal bone mineral density of the proximal femur from clinical CT images. *Arch Osteoporos*. 2022;17:17.

17. Bluemke DA, Moy L, Bredella MA, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the radiology editorial board. *Radiology*. 2020;294: 487–489.

18. Ito M, Ikeda K, Nishiguchi M, et al. Multi-detector row CT imaging of vertebral microstructure for evaluation of fracture risk. *J Bone Miner Res*. 2005;20:1828–1836.

19. Rodríguez-Soto AE, Fritscher KD, Schuler B, et al. Texture analysis, bone mineral density, and cortical thickness of the proximal femur: fracture risk prediction. *J Comput Assist Tomogr*. 2010;34:949–957.

20. Taguchi N, Oda S, Nakaura T, et al. Contrast enhancement in abdominal computed tomography: influence of photon energy of different scanners. *Br J Radiol*. 2018;91:20170285.

21. Mileto A, Guimaraes LS, McCollough CH, et al. State of the art in abdominal CT: the limits of iterative reconstruction algorithms. *Radiology*. 2019;293:491–503.

22. Wagner MW, Namdar K, Biswas A, et al. Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. *Neuroradiology*. 2021;63:1957–1967.

23. Samala RK, Chan HP, Hadjiiski L, et al. Risks of feature leakage and sample size dependencies in deep feature extraction for breast mass classification. *Med Phys*. 2021;48:2827–2837.

24. Gupta P, Cherian KE, Kapoor N, et al. Aortic calcification artifact causing spuriously high bone mineral density in the lumbar spine. *AACE Clin Case Rep*. 2021;7:284–286.

25. Kawashima H, Ichikawa K, Hanaoka S, et al. Relationship between size-specific dose estimates and image quality in computed tomography depending on patient size. *J Appl Clin Med Phys*. 2018;19:246–251.

26. Yasaka K, Akai H, Kunimatsu A, et al. Deep learning with convolutional neural network in radiology. *Jpn J Radiol*. 2018;36:257–272.

27. Ohta Y, Yunaga H, Kitao S, et al. Detection and classification of myocardial delayed enhancement patterns on MR images with deep neural networks: a feasibility study. *Radiol Artif Intell*. 2019; 1:e180061.