Information of Vertebral Bone and Muscle from Computed Tomography Imaging Improves the Prediction Power of Vertebral Fractures Using Deep-Learning Algorithm

Sung Hye Kong, Wonwoo Cho, Sung Bae Park, Jaegul Choo, Jung Hee Kim, Sang Wan Kim, Chan Soo Shin

Submitted to: Journal of Medical Internet Research
on: April 27, 2023

Table of Contents

Information of Vertebral Bone and Muscle from Computed Tomography Imaging Improves the Prediction Power of Vertebral Fractures Using Deep-Learning Algorithm

Sung Hye Kong[1*] MD; Wonwoo Cho[2*] PhD; Sung Bae Park[3] MD; Jaegul Choo[2*] PhD; Jung Hee Kim[4*] MD; Sang Wan Kim[3] MD; Chan Soo Shin[4] MD

[1]Seoul National University Bundang Hospital Seongnam KR
[2]Korea Advanced Institute of Science and Technology Daejeon KR
[3]Seoul National University Boramae Hospital Seoul KR
[4]Seoul National University Hospital Seoul KR
[*]these authors contributed equally

Corresponding Author:
Jung Hee Kim MD
Seoul National University Hospital
101 Dae-hak ro, Jongno gu
Seoul
KR

Abstract

Background: Along with the progressively aging populations, the use of opportunistic computed tomography (CT) scanning is increasing, which could be a valuable method to acquire information on both muscles and bones.

Objective: We aimed to develop and externally validate opportunistic CT-based fracture prediction model using images of bones and muscles by employing a deep learning method.

Methods: The model was developed based on a retrospective longitudinal cohort study of 1,359 patients with abdominal CT images at Seoul National University Bundang Hospital between 2010 and 2019. The model was externally validated in 495 patients from Seoul National University Boramae Hospital. The primary outcome of the study was the incidence of vertebral fracture events within 5 years of follow-up. The image model was developed using an attention convolutional neural network-recurrent neural network model for images of the bone and paravertebral muscles.

Results: The mean age was 72 years, and 69.4% were females. In the development set, the areas under the receiver operator curve (AUROC) for predicting vertebral fractures were superior in images of the bone and paravertebral muscle than in those of the bone-only (0.736±0.003 vs. 0.688±0.001; p<0.001). In the validation cohort, the AUROC for the images of the bone-only and the images of the bone and paravertebral muscle were 0.698±0.001 and 0.729±0.002, respectively; p<0.001). For the clinical models using age, sex, body mass index, use of steroids, smoking status, and secondary osteoporosis, AUROC values for the developmental and validation cohorts were 0.635±0.002 and 0.698±0.021, respectively, significantly lower than those of the image model using bone and muscle (p<0.001). In conclusion, the deep learning model of the convolutional neural network-recurrent neural network structure based on CT images of the muscles and bones could better predict the risk of vertebral fractures than the clinical models.

Conclusions: The model using the images of bone and muscle showed better performance than that using the images of the bone-only. Opportunistic CT screening may contribute to identifying patients with a high fracture risk in the future.

(JMIR Preprints 27/04/2023:48535)
DOI: https://doi.org/10.2196/preprints.48535

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Information of Vertebral Bone and Muscle from Computed Tomography Imaging Improves the Prediction Power of Vertebral Fractures Using Deep-Learning Algorithm

Sung Hye Kong[1,2*], Wonwoo Cho[3,4*], Sung Bae Park[5,6], Jaegul Choo[3,4], Jung Hee Kim[2,7], Sang Wan Kim[2,8], Chan Soo Shin[2,7]
[*]**Authors equally contributed to the study.**

[1]Department of Internal Medicine, Seoul National University Bundang Hospital, Kyonggi,
[2]Department of Internal Medicine, Seoul National University College of Medicine, Seoul,
[3]Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology, Daejeon,
[4]Letsur Inc., Seoul,
[5]Department of Neurosurgery, Seoul National University College of Medicine, Seoul,
[6]Departement of Neurosurgery, Seoul National University Boramae Hospital, Seoul,
[7]Departement of Internal Medicine, Seoul National University Hospital, Seoul,
[8]Departement of Internal Medicine, Seoul National University Boramae Hospital, Seoul, Republic of Korea

***Corresponding authors**
Jung Hee Kim, MD, PhD
Department of Internal Medicine, Seoul National University College of Medicine
101 Dae-hak ro, Jongno gu, Seoul, 03080, Korea
Phone: +82-2-2072-4839. Fax: +82-2-2072-7246
E-mail: jhee1@snu.ac.kr, ORCID: 0000-0003-1932-0234
Jaegul Choo, PhD
Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology,
291 Dae-hak ro, Yuseong-gu, Daejeon, 34141, Korea
Phone: +82-42-350-1813.
E-mail: jchoo@kaist.ac.kr, ORCID: 0000-0003-1071-4835

Abstract

Background: As world ages, the use of opportunistic computed tomography (CT) scanning is increasing, which could be a valuable method to acquire information on both muscles and bones.

Objectives: To develop and externally validate opportunistic CT-based fracture prediction models using images of vertebral bones and paravertebral muscles

Methods: The models were developed based on a retrospective longitudinal cohort study of 1,214 patients with abdominal CT images between 2010 and 2019. The models were externally validated in 495 patients. The primary outcome of this study was defined as the predictive accuracy for identifying vertebral fracture events within 5-year follow-up. The image models were developed using an attention convolutional neural network-recurrent neural network model from images of the vertebral bone and paravertebral muscles.

Results: The mean ages in the development and validation sets were 73 and 68 years, and 69% and 79% of them were female, respectively. The areas under the receiver operator curve (AUROC) for predicting vertebral fractures were superior in images of the vertebral bone and paravertebral muscle than in those of the bone-only in the external validation cohort (0.827, 95% confidence interval [CI] 0.821 - 0.833; 0.815, 95% CI 0.806 - 0.824; p<0.001). The AUROC of these image models were higher than fracture risk assessment (FRAX) model (0.810 for major osteoporotic, 0.780 for hip fracture risk). For the clinical model using age, sex, body mass index, use of steroids, smoking, and possible secondary osteoporosis, type 2 diabetes mellitus, HIV, hepatitis C, and renal failure, AUROC values in the external validation cohort were 0.749 (95% CI 0.736 - 0.762), lower than those of the image model using vertebral bone and muscle (p<0.001).

Conclusion: The model using the images of vertebral bone and muscle showed better performance than that using the images of the bone-only or clinical variables. Opportunistic CT screening may contribute to identifying patients with a high fracture risk in the future.

Keywords: Fracture; Prediction; Deep Learning; Prospective Cohort; Fracture risk assessment.

Introduction

The global aging society has driven an increase in the incidence of fragility fractures and imposed a significant burden on healthcare systems, societies, and, most importantly, on patients and their families [1-3]. Thus, proactively identifying patients having a high risk of fractures is vital. There are well-established methods to evaluate the risk of fractures, such as dual-energy X-ray absorptiometry (DXA) to assess bone mineral density (BMD), which is a reference standard for the diagnosis of osteoporosis [4]. However, a large proportion of patients have never undergone DXA, and 60% of the patients with major osteoporotic fractures do not receive proper treatment to reduce the risk of fractures [5].

Opportunistic computed tomography (CT) scans can be a novel approach for identifying patients having a high risk of fractures. Along with the progressively aging populations, the use of opportunistic CT scanning is increasing with over 80 million examinations performed each year in the US [6]. Retrieval of information that can help assess the fracture risks from opportunistic CT scans does not require additional costs, time, or equipment, and data can be retrospectively acquired. Thus, it may help reduce the efforts associated with screening the patients with high risks of fractures. Several studies have assessed BMD using opportunistic CT scans [7], mainly utilizing the attenuation data of the trabecular bone of the spine [8,9].

There have been significant advances in deep learning techniques for medical image analysis, such as the convolutional neural network (CNN) method [10]. The CNN method facilitates the utilization of highly representative, data-driven image features, arranged in a layered hierarchical structure, which are effective in successfully classifying medical images. Using CNN, various images have been used to classify patients with a high fracture risk [11-14]. Previous studies have primarily focused on the use of X-ray images for fracture detection, with AUROCs reported in the range of 0.73 to 0.80 [14,15]. However, there has been a scarcity of research utilizing CT images and is limited to bone texture analysis. Our study may fill this gap by applying CNN techniques to CT scans, which may provide a more accurate assessment of fracture risk due to the detailed and comprehensive nature of CT imaging. Also, as the paravertebral muscles are among the critical contributing factors to vertebral fractures [16,17], CT could be a valuable method for acquiring information on both the muscle and vertebral bone. Nevertheless, to our knowledge, no previous study has reported using images of both the vertebral bone and muscle from CT scans using the CNN method.

Therefore, we aimed to develop and externally validate a CT-based fracture prediction model using images of vertebral bones and muscles by employing a deep learning method. This study may help identify the patients having high risk of fractures among all the patients who undergo opportunistic CT scans for screening or other purposes.

Methods

Study design and participants

This study was based on a retrospective longitudinal cohort study of 32,435 patients having abdominal CT images at Seoul National University Bundang Hospital (SNUBH) between 2010 and 2019. Patients who met all inclusion criteria were included. Inclusion criteria was as follows: (1) patients who had abdominal CT imaging at SNUBH between 2010 and 2019 and had follow-up images at 5-year timepoint, (2) those who were aged between 50 and 80, (3) those who had followed up over a year. Also, patients who met one of the exclusion criteria were excluded. The exclusion criteria and number of excluded patients were as follows: (1) patients who were younger than 50 or older than 80 (n=2,643), (2) those whose follow-up periods were less than a year (n=8,029), (3) those who had compression fractures or spinal surgery at the baseline (n=3,258) (Figure 1). Thus, 18,505 patients were included in the analysis.

During follow-up, 693 patients experienced vertebral fractures, which remains 17,812 patients who did not. Among 693 patients, after excluding 85 patients owing to the poor image quality or inappropriate CT protocols, 608 patients remained as cases.

For the control group, we selected individuals from the same time frame as the fracture cases. Among 17,812 patients who did not experience fracture, after excluding 2,141 patients with poor image quality, we selected 606 age-, sex-, and body mass index (BMI)-matched individuals at a ratio of 1 patient to 1 control within a similar follow-up period. The fracture events were determined by reviewing medical records, with efforts to exclude any fractures associated with trauma. If patients had multiple CT scans during the follow-up, the earliest CT scan was used.

As a result, 1,214 patients were eligible for analysis and constituted the development set. In addition, we developed an external validation set of 495 patients from Seoul National University Boramae Hospital using the same protocol but without case-control matching, between 2012 and 2013. Developing an external validation set was to assess the performance of the intervention. The use of the same protocol was to ensure consistency in evaluation while allowing for a broader application of the findings in real-world setting.

The study protocol was approved by the Institutional Review Board of the Seoul National University Bundang Hospital (IRB No. B-2104-677-402). The requirement for informed consent was waived owing to the retrospective design of the study. The study was conducted in accordance with the ethical standards laid down in the 1964 World Medical Association Declaration of Helsinki and its later amendments. It also complies with the Ethical Principles for Medical Research.

Primary outcome

The primary outcome of the study was the predictive accuracy of vertebral fracture events between T12 and L4 occurred within five years. Vertebral fractures were defined as morphometric fractures and were confirmed using X-ray or CT images. These images were adjudicated by SHK and JHK, who were blinded to any patient information prior to their assessment. Morphometric vertebral fractures were confirmed by X-ray or reconstructed CT images with measurements of anterior (H_a), middle (H_m), and posterior (H_p) height of each vertebral body from T11 to L4 were measured. Normal population was classified as having no vertebral fracture by gross visual inspection as being within normal range for vertebral height and shape. The mean and standard deviation (SD) of ratio of normal vertebral height were obtained from patients without incident fractures. We first calculated the anterior to posterior (H_a/H_p), middle to posterior (H_m/H_p) and posterior to posterior above and below (H_{pi}/H_{pi+1} and Hpi/H_{pi-1}) ratios. Vertebral fracture was defined if any of the following ratios were more than 3 SDs below the normal mean for that vertebral level, which was 0.91±0.08, as described in previous reports [18,19].

Measurements of clinical factors

Sociodemographic factors, including age, sex, and medical history, were obtained from a

review of electronic medical records at baseline. Height and body weight were measured using standard methods by trained staff using a scale and wall-mounted extensometer and the participants wore lightweight clothes. BMI was calculated as weight divided by height in meters squared (kg/m^2). Current smokers were defined as the patients who were smoking during the study period, while current alcohol consumers were defined as those who consumed three or more units of alcohol daily. The use of glucocorticoids was defined as using oral glucocorticoids or having been exposed to oral glucocorticoids for more than three months at a prednisolone dose >5 mg or its equivalent doses. Possible secondary osteoporosis is defined as osteoporosis that occurs due to factors other than primary menopause or age-related causes. It includes patients with osteoporosis and concurrent diagnosis with type 1 diabetes, osteogenesis imperfecta in adults, hyperthyroidism, hypogonadism, premature menopause (age <45 years), chronic malnutrition, malabsorption, or chronic liver disease [(20)].

CT protocols, image preprocessing, and deep learning techniques

Intravenous contrast-enhanced images were obtained using CT scanners with 64 detector rows (Brilliance; Philips Medical Systems, Cleveland, Ohio, USA). All the patients were placed in a supine position and scanned from the diaphragm to the symphysis pubis. The reference tube current–time product was empirically set, aiming at effective radiation doses of 2 mSv. The effective tube current–time product generally ranges between 25 and 40 mAs. The actual radiation dose was adjusted according to the body size by automatically modulating the tube current (Dose-Right; Philips Medical Systems). The values of tube voltage, collimation, rotation speed, and pitch were 120 kVp, 64 × 0.625 mm, 0.5 second, and 0.891, respectively. Patients were administered 2 mL iopromide/kg (Ultravist 370; Schering, Berlin, Germany) intravenously at a rate of 3 mL/s via the antecubital vein, and scanning was initiated 60 s after the enhancement of the descending aorta reached 150 HU. From each helical scan, the images were reconstructed using a section thickness of 5 mm.

Consecutive image processing was applied to all the CT images for accurate deep learning (DL)-based image analysis. Each axial slice of the abdominal CT scan was resampled to obtain a pixel spacing of 1×1 mm^2. The signal intensity of each CT image was min-max normalized to the (-1–1) range after windowing the Hounsfield unit values in the range of (-200–1000). Subsequently, two classes of image data, vertebral body only (bone-only) and vertebral body with paravertebral muscles (bone+muscle), were extracted from each CT image. Based on the manual annotations of the vertebral body, excluding the intervertebral disc, the vertebral body regions from T12 to L4 were extracted from the CT images, where each bone-only image had 96 x 96 pixels in the axial plane. Centered on the vertebral body, the images of the paravertebral muscle were automatically cropped using a rectangular box (96 x 144 pixels) in the axial plane (Supplementary Figure 1).

DL-based image features for the 5-year risk analysis of vertebral fractures were extracted using the attention convolutional neural network–recurrent neural network (CNN–RNN) model for image data (Figure 2) [(21,22)]. In the CNN–RNN model, ImageNet-pre-trained ResNeXt-50 and gated recurrent units were employed as the CNN encoder backbone and the RNN recurrent decoder, respectively. As inputs of the model, 14 equidistant axial slices were extracted from the T12–L4 vertebrae region of each CT image, where the starting slice was randomly selected at each training iteration for data augmentation. In the training phase, the CNN–RNN model was optimized using the Adam optimizer and cross-entropy loss, where the learning rate and batch size were 1e-5 and 128, respectively. In our image-based fracture prediction model, the utilization of CT images was primarily driven by deep learning methods, particularly CNNs. While specific imaging parameters like attenuation values and density were not directly used as standalone inputs, the CNN's learning process inherently captured these aspects as part of the comprehensive image analysis. The model processed the entire CT images, extracting deep features that potentially included characteristics related to bone and muscle attenuation and density, among others. This approach allowed for a

sophisticated interpretation of the CT scans, identifying nuanced patterns indicative of fracture risk. To incorporate clinical variables into our image-based prediction model, we first standardized clinical variables to ensure consistency and comparability. Following this, we concatenated the standardized variables to the image features in the final layer of the CNN-RNN model.

To understand how the model identifies and differentiates key areas for predicting vertebral fractures in CT images, we employed the gradient-weighted class activation mapping technique (GradCAM). This approach involves highlighting the most crucial regions within the images, marked by a bright red overlay, thereby revealing the model's decision-making process and focal areas for classification (Figure 3). The image models were developed with a high-performance computing server with four NVIDIA GeForce GTX 1080 Ti (NVIDIA, Santa Clara, CA, USA) graphic processing units and the Ubuntu 16.04.4 operating system.

Statistical analyses

For the baseline characteristics, depending on the distribution, continuous parameters are presented as means with standard deviations and categorical data are presented as proportions. Comparisons between the groups having continuous variables were analyzed using Student's *t*-test, whereas χ^2 test was used for the categorical variables. The area under the receiver operating characteristic curve (AUROC) was calculated to compare between the pre-processed images. Cases predicted to have an actual fracture event and experienced it during follow-up were defined as true positive (TP). Those predicted to have, but did not experience, fracture was designated as falsely positive (FP). Cases predicted to be free from fracture events but experienced one during the follow-up were defined as a false negative (FN). True-negative (TN) cases were predicted to be free of fracture events and did not experience fractures during follow-up. Sensitivity and specificity were calculated for each time series as follows: sensitivity= $\frac{TP}{TP+FN}$ and specificity= $\frac{TN}{TN+FP}$. The risk prediction performance measures were gauged using 10-fold cross-validation.

The image-only 5-year risk analyses of vertebral fractures were conducted by applying a fully connected (FC) layer, which generates binary prediction results to the CNN–RNN feature extractor. In addition to the image-only model, clinical models (Models A, B, C, and D) were developed by analyzing the corresponding clinical variables via XGBoost. The clinical variables included age, sex, BMI, use of steroids, smoking status, and possible secondary osteoporosis. Model A included age and sex as independent variables; model B additionally incorporated BMI; model C further included clinical variables such as the use of glucocorticoids, history of alcohol consumption, smoking, and possible secondary osteoporosis; and model D additionally included type 2 diabetes mellitus, HIV , hepatitis C infection [23], and renal failure [24].

The matching of cases and controls in our study was conducted based on age and sex. This process was facilitated using the propensity score matching method, implemented through the MatchIt package in R (version 4.1.2). The R software is developed and maintained by the R Foundation (https://r-project.org). PyTorch and Scikit-learn libraries from Python were used for the analyses. A *P*-value <0.05 was considered significant. Correction for multiple testing was not performed across models. Statistical analyses were performed using Python version 3.8.10 (Python Software Foundation, https://www.python.org). The programs used in the experiments were PyCharm (JetBrains s.r.o., Prague, Czech Republic) and Visual Studio (Microsoft Corp., Redmond, WA, USA).

Results

Clinical characteristics

A total of 1,709 individuals were included in the analysis. The participants were divided into a development set from Seoul National University Bundang Hospital (n=1,214) and an external validation set (n=495) from Seoul National University Boramae Medical Center. As shown in Table 1, The development set had an older average age (72.5 ± 7.9 years) compared to the external validation set (67.6 ± 8.6 years), with a statistically significant difference ($p < 0.001$). The proportion of females was higher in the external validation set (78.8%) than in the development set (69.1%), with this difference also being significant ($p < 0.001$). However, no significant differences were observed in weight and BMI between the two sets. When considering lifestyle factors, there was a higher prevalence of current smokers (23.4% vs. 5.7%, $p < 0.001$) and current drinkers (19.47% vs. 6.9%, $p < 0.001$) in the development set compared to the external validation set. The use of steroids was similar across both groups ($p = 0.564$), while the prevalence of possible secondary osteoporosis was significantly higher in the development set (10.1% vs. 3.2%, $p < 0.001$). During 5 years of follow-up, 454 (37.4%) and 61 (12.3%) individuals experienced vertebral fractures in the development and external validation sets, respectively.

In the development set (n=1,214), participants were matched based on age, sex, and BMI to compare those with incident fractures (n=608) to those without (n=606). There was a less than one-year age difference (mean age 72.0 ± 7.6 years in the non-fracture group vs. 72.9 ± 8.2 years in the fracture group, $p < 0.001$) and a BMI difference of less than 0.5 (mean BMI 23.7 ± 3.4 kg/m² in the non-fracture group vs. 23.5 ± 3.5 kg/m² in the fracture group, $p = 0.391$). Gender distribution was balanced between the two groups (69.64% female in the non-fracture group vs. 68.59% female in the fracture group, $p = 0.957$).

Despite these matched parameters, a higher prevalence of current smokers was noted in the fracture group (27.8% vs. 19.0%, $p = 0.001$), along with a significantly higher use of steroids (18.8% vs. 7.8%, $p < 0.001$) and a greater prevalence of possible secondary osteoporosis (12.7% vs. 7.4%, $p = 0.001$). No significant differences were observed in height, weight, and current drinking status between the two groups. (Table 2).

Comparisons between the performances of image models in predicting vertebral fractures

As demonstrated in Table 3, for the development set, the models using images that included both vertebral bone and paravertebral muscle showed significantly better AUROC, accuracy, and precision values compared to those using bone-only images. Specifically, the bone-only images had an AUROC of 0.677 (0.674 - 0.680) and accuracy of 0.669 (0.665 – 0.673). In contrast, the images including both bone and muscle exhibited an AUROC of 0.739 (0.737 - 0.741) and accuracy of 0.719 (0.715 - 0.722, all $p<0.001$). The FRAX model for major osteoporotic fracture (MOF) and hip fracture showed lower AUROCs of 0.557 and 0.563, respectively, indicating a significantly better performance of our image model (all $p<0.001$).

Similar trends were observed in the external validation set, where bone-only images resulted in an AUROC of 0.815 (0.806 - 0.824), accuracy of 0.754 (0.752 - 0.756), while the combined bone and muscle images demonstrated an AUROC of 0.827 (0.821 - 0.833), accuracy of 0.812 (0.798 - 0.826, all $p<0.001$), though specificity value was similar between the two groups. The FRAX model for MOF and hip fracture had AUROCs of 0.810 and 0.780, respectively. Again, these results confirmed the superior predictive capability of our image-based model (all $p<0.001$).

Comparisons between the performances of images and clinical models

Compared to the clinical models, the image model using vertebral bone and muscle showed significantly higher performance than the clinical models in predicting the vertebral fractures during the five-year follow-up period in the development and external validation sets (Figure 4, Table 4). In

the development set, the images that included vertebral bone and muscle had significantly better AUROC and accuracy than the clinical model D, which included age, sex, BMI, history of alcohol consumption, smoking, possible secondary osteoporosis, type 2 diabetes mellitus, HIV, and hepatitis C infection status, and renal failure. (AUROC, 0.667 [0.661 - 0.672] and accuracy, 0.640 [0.661 - 0.649]; all p<0.001, Table 4). In addition, the performance did not show a significant change when the clinical variables were added to the image-only model (Supplementary Table 1).

As depicted in Figure 4, in the external validation set, the images including vertebral bone and muscle showed a significantly better AUROC and accuracy than the clinical model D (AUROC, 0.749 [0.736 - 0.762] and accuracy, 0.675 [0.643 - 0.707]; all p<0.001). The results were similar for both clinical models A, B, C, and D, which showed poorer performances than that of the image model.

Discussion

In this study, we developed and externally validated a vertebral fracture prediction model by using abdominal CT images. In the developmental cohort, the performance of predicting vertebral fractures represented by AUROC was 0.688 ± 0.001 using images of vertebral bone-only and 0.736 ± 0.003 using images of vertebral bone and paravertebral muscle. In the validation cohort, the performances were 0.698 ± 0.001 and 0.729 ± 0.002 for images of vertebral bone-only and images of vertebral bone and paravertebral muscle, respectively. In addition, the performance of the model using images of vertebral bone and muscle was significantly better than that of the clinical models using age, sex, BMI, use of steroids, smoking status, and possible secondary osteoporosis, which showed performances of 0.635 ± 0.002 and 0.698 ± 0.021, respectively, for the developmental and validation cohorts.

The model showed that the image models using vertebral bone and muscle had a better performance than those using images of vertebral bone-only. It has been reported that osteosarcopenia, a combined occurrence of bone loss and sarcopenia, is one of the critical risk factors for osteoporotic fractures [25,26]. The paravertebral muscles are essential components of the vertebral column and are associated with osteoporotic vertebral fractures [27,28]. In previous studies, information retrieved from muscle images, such as cross-sectional area, volume, and degree of fat infiltration in the paravertebral muscle, was correlated with vertebral stability and the risk of fractures [28,29]. Specifically, Kim *et al.* reported lower cross-sectional areas and greater fat infiltration of the paravertebral muscles in patients with vertebral fractures than in those without fractures [30]. This not only implies the density and quality of the bones that correlate with the risk of fractures but also the quality of the muscles supporting and communicating with the bones [17]. Fat infiltration in the muscles, called myosteatosis, has been reported to be associated with an increased risk of fractures [17,31]. Thus, in line with previous studies, the study results imply that information from the images of the paravertebral muscles, in addition to the information from the images of vertebral bones, can help predict vertebral fractures more accurately.

In this study, the image-based learning model with the images of both vertebral bone and muscle showed better performance than that of the clinical variable-based models. This is consistent with a previous report that information from the images of vertebral bones and muscles from CT scans can be used to predict major osteoporotic fractures, and is comparable with FRAX [32]. Another group reported different algorithms using opportunistic CT-based bone assessments for osteoporotic fracture prediction [33]. They showed that CT-based predictors (vertebral compression fractures, simulated DXA T-scores, and lumbar trabecular density) with metadata of age and sex showed better performance in AUROC than FRAX [33]. However, in this model, muscle information was not considered [33], which may further improve the performance. In addition to the attenuation information, we used information from the image itself, which harnesses information on the quality of the bone and muscle structure, similar to the trabecular bone score [13]. The trabecular bone score is an algorithm used to calculate the microstructure of the bone based on DXA images [34]. In addition, more than 50% of the osteoporotic fractures occur in the patients with a normal or osteopenic range of BMD [35], which implies that the microarchitecture of the bone is also a key determinant of bone strength [36]. Similarly, in our study, the model used the information on the qualities of bones and muscles from CT images, demonstrating the potential value of CT images that may include rich and various informative data for the metabolic diseases of bones and muscles.

We also observed that the performance did not significantly change when clinical variables were added to the image-only model. There is a possibility that information such as age and gender could already be reflected to some extent in the image itself [37]. Therefore, there could be an insignificant improvement in the performance because the information poses a redundant input to the model. It is widely accepted that there is a noticeable sex difference in the size of the vertebral body and paravertebral muscles [37], and body mass index could be positively correlated with the size of the vertebrae and muscles. In addition, although it was based on high-resolution peripheral quantitative

CT, each bone has different characteristics according to the age and sex, such as calcification and size, which could have influenced the current analysis [38]. In addition, the vertebral endplate calcification increases with age, implying that age information can be reflected in the image [39]. In addition to previous studies, smoking and alcohol consumption status can be associated with low muscle mass [40], which may explain why adding simple clinical variables to the image may not significantly improve the model, as the image already contains some clinical information. The results are clinically promising, as they could be utilized in the future, as only opportunistic CT scans without detailed clinical variables may automatically provide the risk of osteoporotic fractures.

To extract pertinent information from each CT scan, we designed an image-only model to prevent overfitting and to focus on the essential regions. Since 3-dimensional (3D) CNN models, which have a large number of parameters to be optimized, tend to overfit the training data [41], the CNN encoder of our model took consecutive 2D images as its input data while keeping their sequential information with the RNN decoder [42]. The input processing strategy served as a robust data augmentation method because our model could exploit different 2D image sets from a single 3D CT scan at each training iteration. In addition, an attention module was applied to the CNN encoder to further enhance its robustness. The attention module automatically guided the image model to concentrate on essential regions [43] for the prediction of vertebral fractures. Thus, the attention CNN–RNN model avoids making predictions based on background regions, except for the vertebral body and paravertebral muscles. Unlike previous CNN models-based deep learning algorithms, which were limited to 2D X-ray analysis or bone texture analysis, our CNN–RNN model showed robust performance in fracture prediction. Owing to its design to mitigate the overfitting problem of conventional 3D CNN models [42], the CNN–RNN model could extract effective information from 3D CT images, which was intractable in previous approaches. In addition, the attention module forced our model to focus on important regions in the CT images by removing the effects of the background regions [43].

The study had several limitations. The dataset did not contain BMD due to the retrospective study design, which is an essential predictor for osteoporotic fracture. It was difficult to compare the clinical model containing BMD with the image model. The model showed a 5-year fracture prediction model, instead of a 10-year model, owing to the follow-up duration of the dataset, which is relatively short to be utilized in real-world practice. Thus, due to the short time frame, we could not show the results for nonvertebral fractures because the number of cases was too small. In addition, the paravertebral muscles were included without distinction among the psoas, intervertebralis, multifidus, longissimus, iliocostalis, and quadratus lumborum muscles. Therefore, it is difficult to interpret the contribution of each muscle. In addition, the number of images in the development set may not be sufficient for model optimization. Moreover, the utilization could be low in various contrast settings because it was based on contrast CT scans. There was also the disparity in vertebral fracture incidence between the development and the external validation set, which may affect the external validity and generalizability of our fracture prediction model. The retrospective nature inherently carries the potential for selection bias, including confounding by indication. While we have employed propensity score adjustment to mitigate this bias, it is important to acknowledge that residual bias may still be present. Another limitation is the exclusion of x-ray imaging data with poor quality from our models. This decision might have introduced detection bias, that it may have impacted the diagnostic accuracy of our models in correctly identifying positive versus negative fracture cases. Also, we could not assess the reproducibility of these measurements through inter-examiner and intra-examiner kappa value assessments, which could be considered a limitation of our study. Future prospective studies could benefit from including such reproducibility assessments.

This study has several strengths. This study was longitudinally designed to observe future fracture events in patients who did not have baseline fractures. Furthermore, in the developmental cohort, we used controls with matched clinical variables, which made it possible to attenuate the effects of major clinical variables in the model. It was also externally validated, which helped prove the

generalizability of the model. In addition, the model used the image itself as an input, which made it possible to utilize the information on vertebral bone and muscle quality and quantity. This inclusion of muscle image reflected the interplay between muscle health and fracture risk. For instance, factors like muscle mass and muscle steatosis, which is visible in CT images as darker and more heterogeneous areas compared to normal muscle, could be crucial inputs. These muscle attributes, automatically analyzed by the CNN, contribute significantly to the model's ability to discern patients at higher risk of fractures, offering a more comprehensive view than bone analysis alone. In addition, by sequentially applying bones and muscles to the model, it was possible to check the degree of contribution of muscles and bones to the model performance, thereby increasing the interpretability of the model. In addition, the differences of clinical characteristics between development and external validation sets were purposefully leveraged to assess the generalizability of our model across populations with varying clinical profiles.

In this study, we showed that a deep–learning model of the CNN–RNN structure based on CT images of the muscle and vertebral bone could help predict the risk of vertebral fractures. The model using images of the vertebral bone and muscle showed better performance than the model using images of the vertebral bone-only. This implies that the information from the muscle images provides additional key information for predicting fractures. In addition, the model using images showed better performance than the model using clinical variables, suggesting that the images can provide useful information in addition to having known clinical variables. This study has clinical significance in suggesting that opportunistic CT screening with deep-learning algorithms utilizing bone and muscle may also contribute to identifying patients with a high fracture risk in the future. Further prospective studies are needed to broaden the applicability of the model.

Acknowledgements
Funding
The study was funded by the National Research Foundation of Korea (Grant number 2020R1A2C2011587, 2021R1A2C2003410).

Conflict of interest
No potential conflict of interest relevant to this article was reported.

Data Availability Statement
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. Tran O, Silverman S, Xu X, Bonafede M, Fox K, McDermott M, et al. Long-term direct and indirect economic burden associated with osteoporotic fracture in US postmenopausal women. Osteoporos Int. Jun 2021;32(6):1195-205. Epub 2021/01/08.
2. Williams SA, Daigle SG, Weiss R, Wang Y, Arora T, Curtis JR. Economic Burden of Osteoporosis-Related Fractures in the US Medicare Population. Ann Pharmacother. Jul 2021;55(7):821-9. Epub 2020/11/06.
3. Ahn SH, Park SM, Park SY, Yoo JI, Jung HS, Nho JH, et al. Osteoporosis and Osteoporotic Fracture Fact Sheet in Korea. J Bone Metab. Nov 2020;27(4):281-90. Epub 2020/12/16.
4. Stone KL, Seeley DG, Lui LY, Cauley JA, Ensrud K, Browner WS, et al. BMD at multiple sites and risk of fracture of multiple types: long-term results from the Study of Osteoporotic Fractures. J Bone Miner Res. Nov 2003;18(11):1947-54. Epub 2003/11/11.
5. Amarnath AL, Franks P, Robbins JA, Xing G, Fenton JJ. Underuse and Overuse of Osteoporosis Screening in a Regional Health System: a Retrospective Cohort Study. J Gen Intern Med. Dec 2015;30(12):1733-40. Epub 2015/05/20.
6. IMV. IMV 2022 CT Market Outlook Report. Des Plaines, IL: IMV Medical Information Division. 2022.
7. Lenchik L, Weaver AA, Ward RJ, Boone JM, Boutin RD. Opportunistic Screening for Osteoporosis Using Computed Tomography: State of the Art and Argument for Paradigm Shift. Current Rheumatology Reports. 2018/10/13 2018;20(12):74.
8. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. Annals of internal medicine. 2013;158(8):588-95.
9. Alacreu E, Moratal D, Arana E. Opportunistic screening for osteoporosis by routine CT in Southern Europe. Osteoporosis international. 2017;28(3):983-90.
10. Dutta P, Upadhyay P, De M, Khalkar R. Medical image analysis using deep convolutional

neural networks: CNN architectures and transfer learning. 2020 International Conference on Inventive Computation Technologies (ICICT): IEEE; 2020. p. 175-80.

11. Kong SH, Lee JW, Bae BU, Sung JK, Jung KH, Kim JH, et al. Development of a Spine X-Ray-Based Fracture Prediction Model Using a Deep Learning Algorithm. Endocrinol Metab (Seoul). Aug 2022;37(4):674-83. Epub 2022/08/05.
12. Derkatch S, Kirby C, Kimelman D, Jozani MJ, Davidson JM, Leslie WD. Identification of Vertebral Fractures by Convolutional Neural Networks to Predict Nonvertebral and Hip Fractures: A Registry-based Cohort Study of Dual X-ray Absorptiometry. Radiology. Nov 2019;293(2):405-11. Epub 2019/09/19.
13. Kong SH, Hong N, Kim JW, Kim DY, Kim JH. Application of the Trabecular Bone Score in Clinical Practice. J Bone Metab. May 2021;28(2):101-13. Epub 2021/06/16.
14. Kong SH, Shin CS. Applications of Machine Learning in Bone and Mineral Research. Endocrinol Metab (Seoul). Oct 2021;36(5):928-37. Epub 2021/10/23.
15. Muehlematter UJ, Mannil M, Becker AS, Vokinger KN, Finkenstaedt T, Osterhoff G, et al. Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning. Eur Radiol. May 2019;29(5):2207-17. Epub 2018/12/07.
16. Zhang SB, Chen H, Xu HW, Yi YY, Fang XY, Wang SJ. Computed tomography-based paravertebral muscle density predicts subsequent vertebral fracture risks independently of bone mineral density in postmenopausal women following percutaneous vertebral augmentation. Aging Clin Exp Res. Nov 2022;34(11):2797-805. Epub 2022/08/25.
17. Kim HK, Kim CH. Quality Matters as Much as Quantity of Skeletal Muscle: Clinical Implications of Myosteatosis in Cardiometabolic Health. Endocrinol Metab (Seoul). Dec 2021;36(6):1161-74. Epub 2022/01/06.
18. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a

semiquantitative technique. J Bone Miner Res. Sep 1993;8(9):1137-48. Epub 1993/09/01.

19. Shin CS, Kim MJ, Shim SM, Kim JT, Yu SH, Koo BK, et al. The prevalence and risk factors of vertebral fractures in Korea. J Bone Miner Metab. Mar 2012;30(2):183-92. Epub 2011/07/21.

20. Kanis JA, Johansson H, McCloskey EV, Liu E, Åkesson KE, Anderson FA, et al. Previous fracture and subsequent fracture risk: a meta-analysis to update FRAX. Osteoporos Int. Dec 2023;34(12):2027-45. Epub 2023/08/11.

21. Wang J, Yang Y, Mao J, Huang Z, Huang C, Xu W. Cnn-rnn: A unified framework for multi-label image classification. Proceedings of the IEEE conference on computer vision and pattern recognition2016. p. 2285-94.

22. Khaki S, Wang L, Archontoulis SV. A cnn-rnn framework for crop yield prediction. Frontiers in Plant Science. 2020;10:1750.

23. Dong HV, Cortés YI, Shiau S, Yin MT. Osteoporosis and fractures in HIV/hepatitis C virus coinfection: a systematic review and meta-analysis. Aids. Sep 10 2014;28(14):2119-31. Epub 2014/07/01.

24. Ganesan K, Jandu JS, Anastasopoulou C, Ahsun S, Roane D. Secondary Osteoporosis. StatPearls. Treasure Island (FL) ineligible companies. Disclosure: Jagmohan Jandu declares no relevant financial relationships with ineligible companies. Disclosure: Catherine Anastasopoulou declares no relevant financial relationships with ineligible companies. Disclosure: Sana Ahsun declares no relevant financial relationships with ineligible companies. Disclosure: Douglas Roane declares no relevant financial relationships with ineligible companies.: StatPearls Publishing

25. Teng Z, Zhu Y, Teng Y, Long Q, Hao Q, Yu X, et al. The analysis of osteosarcopenia as a risk factor for fractures, mortality, and falls. Osteoporos Int. Nov 2021;32(11):2173-83. Epub

2021/04/21.

26. Cedeno-Veloz B, López-Dóriga Bonnardeauxa P, Duque G. [Osteosarcopenia: A narrative review]. Rev Esp Geriatr Gerontol. Mar-Apr 2019;54(2):103-8. Epub 2018/11/26.

27. Yagi M, Hosogane N, Watanabe K, Asazuma T, Matsumoto M. The paravertebral muscle and psoas for the maintenance of global spinal alignment in patient with degenerative lumbar scoliosis. Spine J. Apr 2016;16(4):451-8. Epub 2015/07/15.

28. Habibi H, Takahashi S, Hoshino M, Takayama K, Sasaoka R, Tsujio T, et al. Impact of paravertebral muscle in thoracolumbar and lower lumbar regions on outcomes following osteoporotic vertebral fracture: a multicenter cohort study. Arch Osteoporos. Jan 3 2021;16(1):2. Epub 2021/01/04.

29. Choi MK, Kim SB, Park CK, Malla HP, Kim SM. Cross-Sectional Area of the Lumbar Spine Trunk Muscle and Posterior Lumbar Interbody Fusion Rate: A Retrospective Study. Clin Spine Surg. Jul 2017;30(6):E798-e803. Epub 2016/09/14.

30. Kim JY, Chae SU, Kim GD, Cha MS. Changes of paraspinal muscles in postmenopausal osteoporotic spinal compression fractures: magnetic resonance imaging study. J Bone Metab. Nov 2013;20(2):75-81. Epub 2014/02/14.

31. Lang T, Cauley JA, Tylavsky F, Bauer D, Cummings S, Harris TB. Computed tomographic measurements of thigh muscle cross-sectional area and attenuation coefficient predict hip fracture: the health, aging, and body composition study. J Bone Miner Res. Mar 2010;25(3):513-9. Epub 2010/04/28.

32. Pickhardt PJ, Graffy PM, Zea R, Lee SJ, Liu J, Sandfort V, et al. Automated Abdominal CT Imaging Biomarkers for Opportunistic Prediction of Future Major Osteoporotic Fractures in Asymptomatic Adults. Radiology. 2020/10/01 2020;297(1):64-72.

33. Dagan N, Elnekave E, Barda N, Bregman-Amitai O, Bar A, Orlovsky M, et al. Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid

in FRAX underutilization. Nat Med. Jan 2020;26(1):77-82. Epub 2020/01/15.

34. Pothuaud L, Carceller P, Hans D. Correlations between grey-level variations in 2D projection images (TBS) and 3D microarchitecture: applications in the study of human trabecular bone microarchitecture. Bone. Apr 2008;42(4):775-87. Epub 2008/02/01.

35. Siris ES, Chen YT, Abbott TA, Barrett-Connor E, Miller PD, Wehren LE, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. Arch Intern Med. May 24 2004;164(10):1108-12. Epub 2004/05/26.

36. Silva BC, Leslie WD, Resch H, Lamy O, Lesnyak O, Binkley N, et al. Trabecular bone score: a noninvasive analytical method based upon the DXA image. J Bone Miner Res. Mar 2014;29(3):518-30. Epub 2014/01/21.

37. Gilsanz V, Boechat MI, Gilsanz R, Loro ML, Roe TF, Goodman WG. Gender differences in vertebral sizes in adults: biomechanical implications. Radiology. 1994;190(3):678-82.

38. Kazakia GJ, Nirody JA, Bernstein G, Sode M, Burghardt AJ, Majumdar S. Age-and gender-related differences in cortical geometry and microstructure: Improved sensitivity by regional analysis. Bone. 2013;52(2):623-31.

39. Bernick S, Cailliet R. Vertebral end-plate changes with aging of human vertebrae. Spine. 1982;7(2):97-102.

40. Zhong J, Xie W, Wang X, Dong X, Mo Y, Liu D, et al. The Prevalence of Sarcopenia among Hunan Province Community-Dwelling Adults Aged 60 Years and Older and Its Relationship with Lifestyle: Diagnostic Criteria from the Asian Working Group for Sarcopenia 2019 Update. Medicina. 2022;58(11):1562.

41. Kumawat S, Raman S. Lp-3dcnn: Unveiling local phase in 3d convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition2019. p. 4903-12.

42. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural

networks on sequence modeling. arXiv preprint arXiv:14123555. 2014.

43. Jetley S, Lord NA, Lee N, Torr PH. Learn to pay attention. arXiv preprint arXiv:180402391. 2018.

Table 1. Clinical Characteristics of Development and External Validation Sets

	Development set (n=1,214)	External validation set (n=495)	*p*
Age, years	72.5 ± 7.9	67.6 ± 8.6	< 0.001
Female	839 (69.1%)	390 (78.8%)	< 0.001
Height, cm	157.0 ± 8.4	155.2 ± 7.9	< 0.001
Weight, kg	58.2 ± 9.9	57.6 ± 9.7	0.218
BMI, kg/m^2	23.6 ± 3.5	23.9 ± 3.6	0.175
Current smoker	284 (23.4%)	28 (5.7%)	<0.001
Current drinker	236 (19.4%)	34 (6.9%)	<0.001
Use of steroids	161 (13.3%)	57 (11.5%)	0.564
Possible secondary osteoporosis	122 (10.1%)	16 (3.2%)	<0.001
Vertebral fracture during 5 years	608 (50.0%)	61 (12.3%)	<0.001

BMI, body mass index. Use of steroids was defined as the use of prednisolone 5 mg daily or equivalent over 3 months. Possible secondary osteoporosis includes type 1 diabetes, osteogenesis imperfecta in adults, hyperthyroidism, hypogonadism, premature menopause (<45 years), chronic malnutrition, malabsorption, and chronic liver disease. Numbers are presented as numbers (percentages) or mean (standard deviation). The variables between groups were compared using the Student t-test for continuous variables and the $\chi 2$ test for categorical variables.

Table 2. Baseline Clinical Characteristics of Development Set Comparing Incident Fracture Groups

	Incident fracture (-) (n=606)	Incident fracture (+) (n=608)	*p*
Age, years	72.0 ± 7.6	72.9 ± 8.2	<0.001
Female, n (%)	422 (69.6%)	417 (68.6%)	0.957
Height, cm	156.9 ± 8.5	157.1 ± 8.3	0.319
Weight, kg	58.29 ± 9.47	58.11 ± 10.4	0.234
BMI, kg/m^2	23.7 ± 3.4	23.5 ± 3.5	0.391
Current smoker, n (%)	115 (19.0%)	169 (27.8%)	0.001
Current drinker, n (%)	113 (18.7%)	123 (20.2%)	0.923
Use of steroids, n (%)	47 (7.8%)	114 (18.8%)	<0.001
Possible secondary osteoporosis, n (%)	45 (7.4%)	77 (12.7%)	0.001

Fracture (-) and (+) groups represent participants who did not and did experience fractures at 5 years of follow-up, respectively. BMI, body mass index. Use of steroids was defined as the use of prednisolone 5 mg daily or equivalent over 3 months. Possible secondary osteoporosis includes type 1 diabetes, osteogenesis imperfecta in adults, hyperthyroidism, hypogonadism, or premature menopause (<45 years), chronic malnutrition, malabsorption, and chronic liver disease. Numbers are presented as numbers (percentages) or mean (standard deviation). The variables between groups were compared using the Student t-test for continuous variables and the $\chi 2$ test for categorical variables.

Table 3. Performance Comparisons of Image Models in Predicting Vertebral Fractures

	Development set			*External validation set*		
	Bone-only	**Bone + Muscle**	***P***	**Bone-only**	**Bone + Muscle**	***P***
AUROC	0.677 (0.674 - 0.680)	0.739 (0.737 - 0.741)	<0.001	0.815 (0.806 - 0.824)	0.827 (0.821 - 0.833)	0.042
Accuracy	0.669 (0.665 – 0.673)	0.719 (0.715 - 0.722)	<0.001	0.754 (0.752 - 0.756)	0.812 (0.798 - 0.826)	<0.001
Sensitivity	0.746 (0.739 - 0.753)	0.761 (0.746 - 0.776)	0.230	0.645 (0.613 - 0.677)	0.704 (0.675 - 0.733)	0.054
Specificity	0.601 (0.586 - 0.616)	0.634 (0.625 - 0.643)	0.002	0.844 (0.810 - 0.877)	0.855 (0.835 - 0.875)	0.427

AUROC, the area under the receiver operator curve.

Table 4. Performance Comparisons of Image and Clinical Variable Models in Predicting Fractures

	AUROC	*p*	Accuracy	*p*	Sensitivity	*p*	Specificity	*p*
Development set								
Image-only	0.739 (0.737 - 0.741)	-	0.719 (0.716 - 0.722)	-	0.761 ± 0.024 (0.746 - 0.776)	-	0.634 (0.625 - 0.643)	-
Clinical model A	0.647 (0.643 - 0.651)	<0.001	0.620 (0.614 - 0.626)	<0.001	0.681 (0.643 - 0.719)	0.025	0.575 (0.549 - 0.601)	<0.001
Clinical model B	0.631 (0.626 - 0.636)	<0.001	0.612 (0.610 - 0.614)	<0.001	0.675 (0.639 - 0.711)	0.023	0.558 (0.517 - 0.598)	0.003
Clinical model C	0.663 (0.659 - 0.667)	<0.001	0.637 (0.631 - 0.643)	<0.001	0.723 (0.694 - 0.752)	0.114	0.553 (0.521 - 0.585)	<0.001
Clinical model D	0.667 (0.661 - 0.672)	<0.001	0.640 (0.661 - 0.649)	<0.001	0.729 (0.690 - 0.768)	0.129	0.560 (0.527 - 0.593)	0.005
FRAX (MOF)	0.557	<0.001	0.557	<0.001	0.442	<0.001	0.672	<0.001
FRAX (hip)	0.563	<0.001	0.556	<0.001	0.449	<0.001	0.663	<0.001
External validation set								
Image-only	0.827 (0.821 - 0.833)	-	0.812 (0.798 - 0.826)	-	0.704 (0.675 - 0.733)	-	0.855 (0.834 - 0.875)	-
Clinical model A	0.731 (0.725 - 0.737)	<0.001	0.651 (0.629 - 0.673)	<0.001	0.715 (0.683 - 0.747)	<0.001	0.656 (0.629 - 0.683)	<0.001
Clinical model B	0.733 (0.725 - 0.737)	<0.001	0.654 (0.625 - 0.683)	<0.001	0.728 (0.673 - 0.783)	<0.001	0.662 (0.621 - 0.703)	<0.001
Clinical model C	0.745 (0.733 - 0.757)	<0.001	0.669 (0.646 - 0.692)	<0.001	0.713 (0.678 - 0.748)	<0.001	0.720 (0.689 - 0.751)	<0.001
Clinical model D	0.749 (0.736 - 0.762)	<0.001	0.675 (0.643 - 0.707)	<0.001	0.729 (0.690 - 0.768)	<0.001	0.686 (0.650 - 0.722)	<0.001
FRAX (MOF)	0.810	<0.001	0.810	<0.001	0.262	<0.001	0.887	<0.001
FRAX (hip)	0.780	<0.001	0.685	<0.001	0.705	<0.001	0.682	<0.001

Numbers are presented as mean ± standard deviation. AUROC, the area under the receiver operator curve; MOF, major osteoporotic fracture. Image model represents the model using bone and muscle. Model A includes age and sex, model B additionally includes body mass index, and model C additionally includes history of drinking, smoking, and possible secondary osteoporosis.

Supplement Table 1. Performance Comparisons of Image and Combination Model of Image and Clinical Variables

	AUROC	*p*	Accuracy	*p*	Sensitivity	*p*	Specificity	*p*
Development set								
Image-only	0.739 (0.737 - 0.741)	-	0.719 (0.716 - 0.722)	-	0.761 (0.746 - 0.776)	-	0.634 (0.625 - 0.643)	-
Image + model A	0.734 (0.729 - 0.739)	0.417	0.708 (0.704 - 0.712)	0.001	0.750 (0.730 - 0.770)	0.562	0.642 (0.623 - 0.661)	0.766
Image + model B	0.732 (0.726 - 0.738)	0.251	0.715 (0.713 - 0.717)	0.424	0.755 (0.738 - 0.772)	0.764	0.629 (0.615 - 0.642)	0.452
Image + model C	0.736 (0.734 - 0.738)	0.108	0.720 (0.719 - 0.721)	0.198	0.759 (0.745 - 0.773)	0.361	0.630 (0.613 - 0.647)	0.144
Image + model D	0.737 (0.733 - 0.741)	0.077	0.718 (0.716 - 0.720)	0.259	0.762 (0.750 - 0.774)	0.391	0.632 (0.621 - 0.643)	0.047
External validation set								
Image-only	0.827 (0.821 - 0.833)	-	0.812 (0.798 - 0.826)	-	0.704 (0.675 - 0.733)	-	0.855 (0.834 - 0.875)	-
Image + model A	0.810 (0.803 - 0.816)	<0.001	0.788 (0.770 - 0.806)	0.015	0.620 (0.595 - 0.645)	<0.001	0.795 (0.778 - 0.812)	<0.001
Image + model B	0.820 (0.812 - 0.828)	0.009	0.802 (0.775 - 0.829)	0.138	0.681 (0.649 - 0.713)	0.215	0.810 (0.783 - 0.837)	0.002
Image + model C	0.825 (0.820 - 0.830)	0.407	0.805 (0.767 - 0.843)	0.378	0.701 (0.681 - 0.721)	0.477	0.840 (0.816 - 0.864)	0.003
Image + model D	0.824 (0.817 - 0.831)	0.315	0.813 (0.796 - 0.830)	0.726	0.699 (0.680 - 0.718)	0.574	0.849 (0.817 - 0.881)	0.039

AUROC, the area under the receiver operator curve. None of them were significant compared to image-only model. Model A includes age and sex, model B additionally includes body mass index, model C additionally includes the use of steroid, history of drinking, smoking, and possible secondary osteoporosis, and model D additionally includes type 2 diabetes mellitus, HIV, and hepatitis C infection status, and renal failure.

Figure Legends

Figure 1. Flowchart of Participant Selection for Study on Vertebral Fracture Prediction

CT, computed tomography; BMI, body mass index.

Figure 2. Architecture of CNN-RNN Model for Vertebral Fracture Prediction

CNN, convolutional neural network; RNN, recurrent neural network; FC, fully connected.

Figure 3. GradCAM Heatmaps of Image Models Using (A) Bone-only and (B) Bone+Muscle Images

A heatmap was generated by GradCAM (Gradient-weighted Class Activation Mapping), a technique used to visualize which parts of an input image are important for a neural network's decision, especially in Convolutional Neural Networks.

Figure 4. Receiver Operator Curves Comparing Models for 5-year Vertebral Fracture Prediction

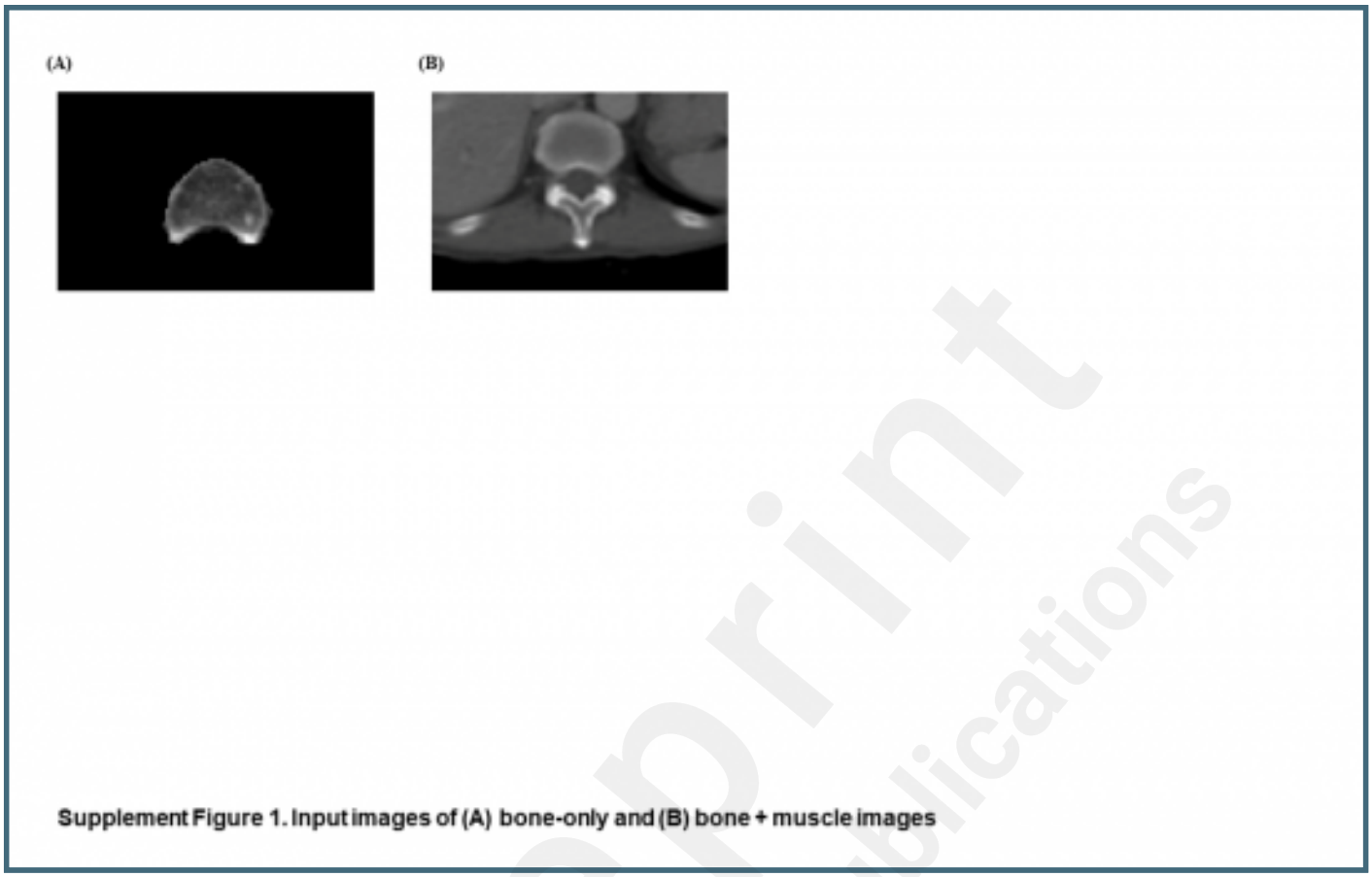
Model A includes age and sex, model B additionally includes body mass index, and model C additionally includes a history of drinking, smoking, and possible secondary osteoporosis.

Supplement Figure 1. Input Images for Bone-only and Bone+Muscle Models

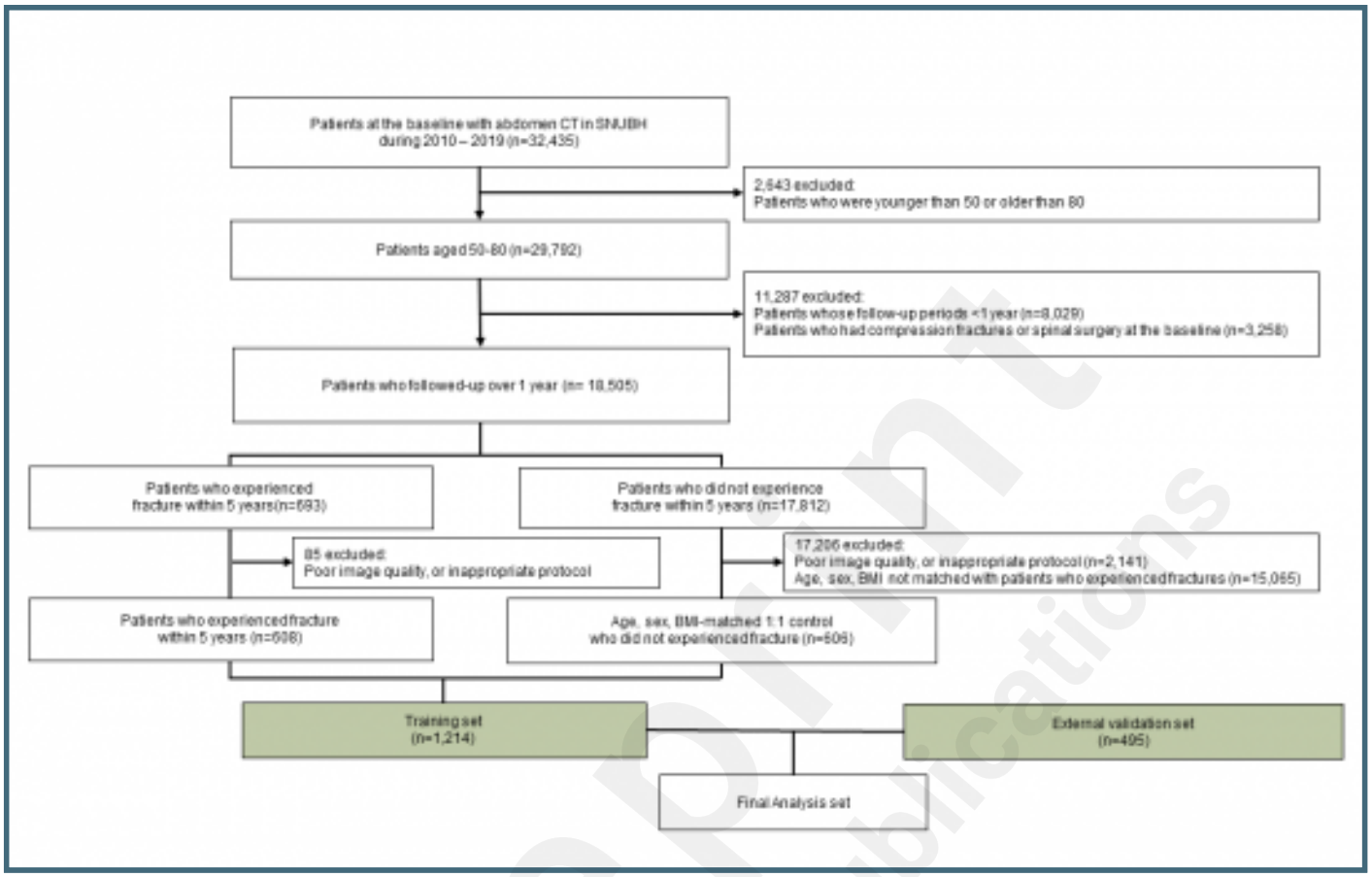
Supplementary Files

Figures

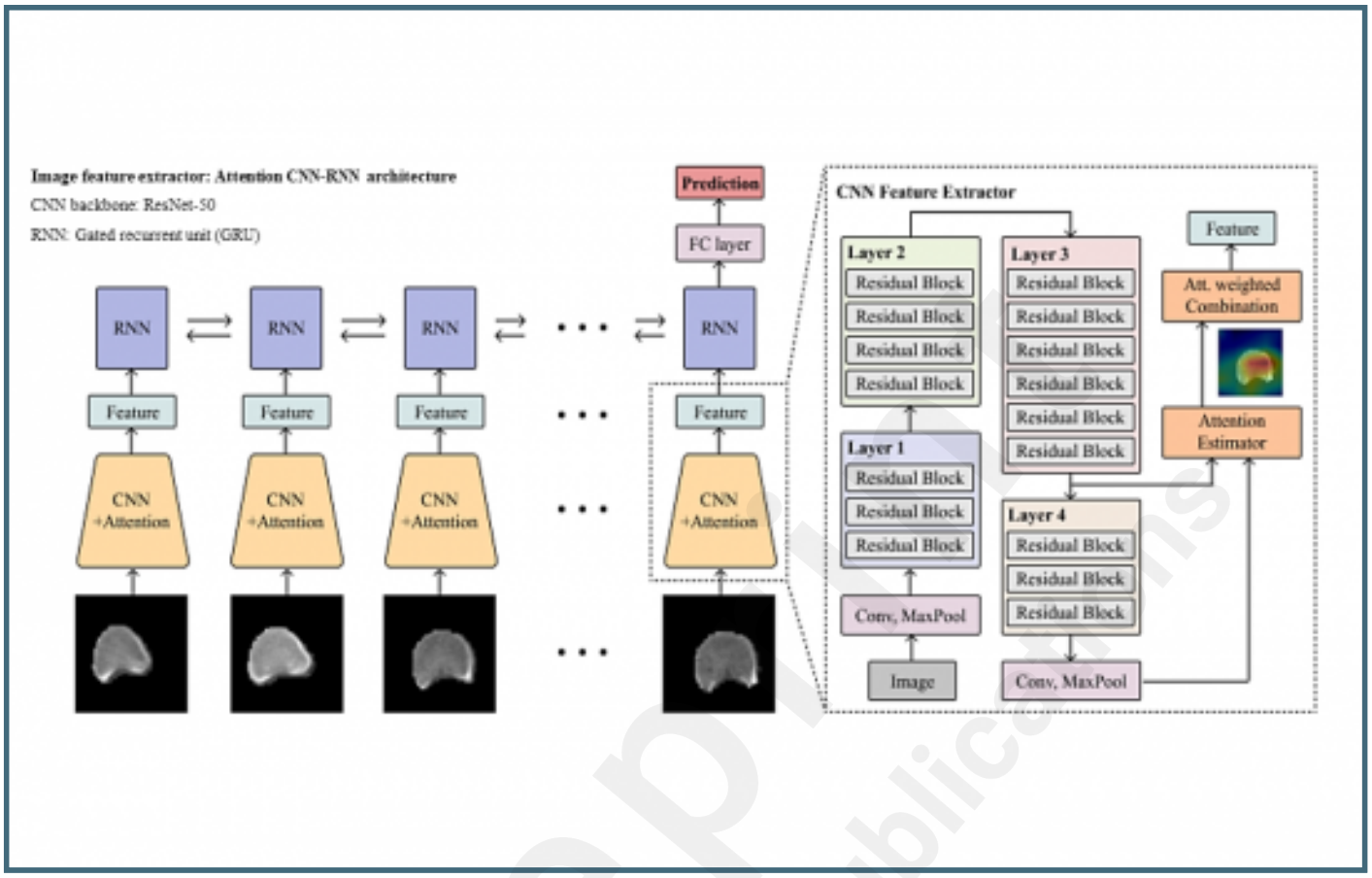
Supplement Figure 1. Input images of (A) bone-only and (B) bone + muscle images.



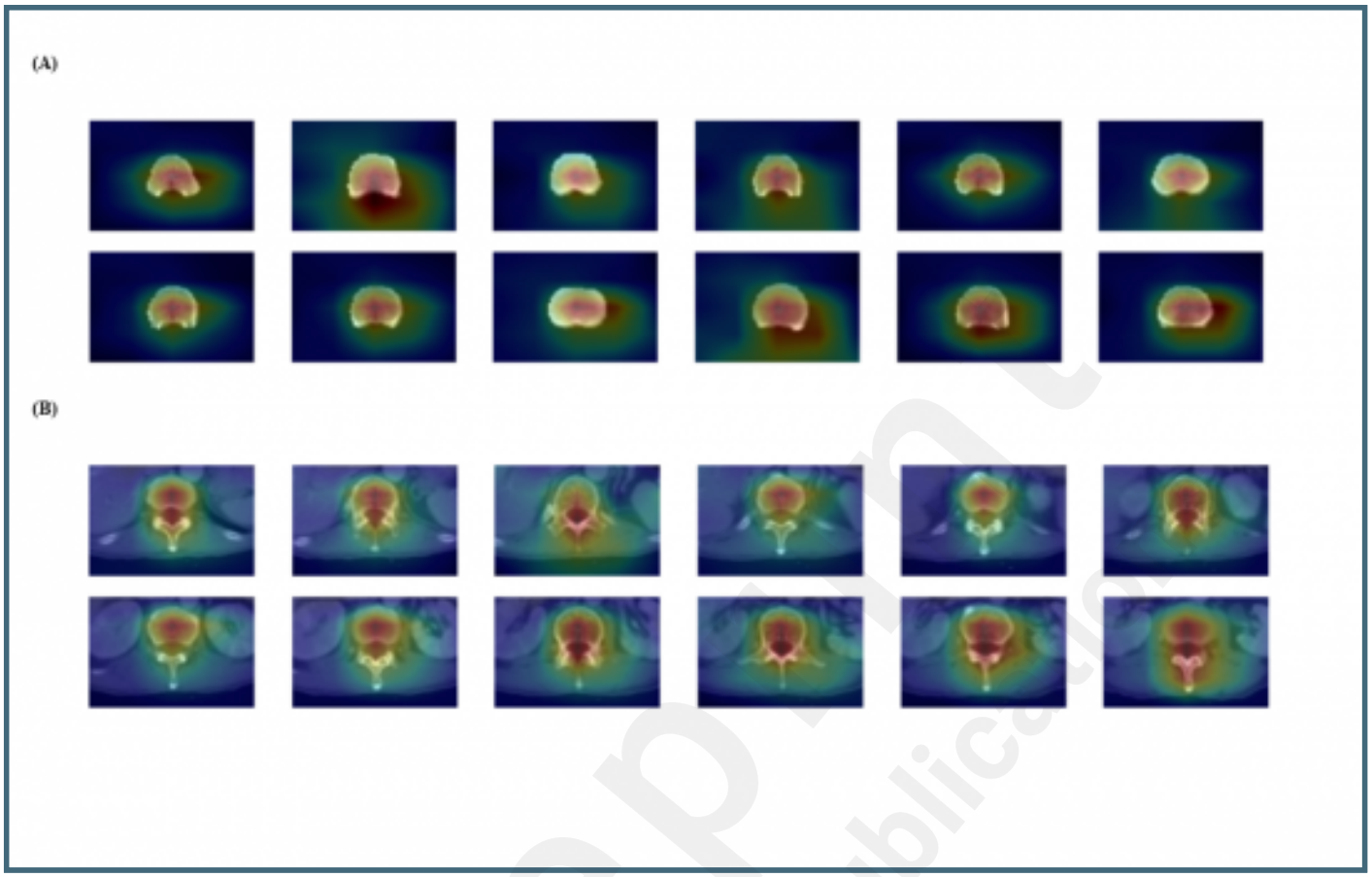
Selection of study participants.



The architecture of image analysis.



Heatmap of image models using (A) bone-only and (B) bone + muscle images.



Receiver operator curves of models for 5-year fracture prediction comparing models using images to models using clinical variables.