



# Femoral osteoporosis prediction model using autosegmentation and machine learning analysis with PyRadiomics on abdomen-pelvic computed tomography (CT)

Min Su Park<sup>1^</sup>, Hong Il Ha<sup>1^</sup>, Hyun Kyung Lim<sup>2^</sup>, Junhee Han<sup>3^</sup>, Seongyong Pak<sup>4^</sup>

<sup>1</sup>Department of Radiology, Hallym University Sacred Heart Hospital, Anyang-si, Gyeonggi-do, Republic of Korea; <sup>2</sup>Department of Radiology, Soonchunhyang University Seoul Hospital, Seoul, Republic of Korea; <sup>3</sup>Department of Statistics and Data Science Convergence Research Center, Hallym University, Chuncheon-si, Gangwon-do, Republic of Korea; <sup>4</sup>CT Research Collaboration, Siemens-Healthineers, Seoul, Republic of Korea

*Contributions:* (I) Conception and design: HI Ha; (II) Administrative support: HI Ha; (III) Provision of study materials or patients: HI Ha, HK Lim, J Han; (IV) Collection and assembly of data: MS Park, HI Ha, HK Lim; (V) Data analysis and interpretation: MS Park, HI Ha, HK Lim; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Hong Il Ha, MD, PhD. Department of Radiology, Hallym University Sacred Heart Hospital, 22 Gwanpyeong-ro 170beon-gil, Dongan-gu, Anyang 14068, Korea. Email: ha.hongil@gmail.com.

**Background:** With the advancement of artificial intelligence technology and radiomics analysis, opportunistic prediction of osteoporosis with computed tomography (CT) is a new paradigm in osteoporosis screening. This study aimed to assess the diagnostic performance of osteoporosis prediction by the combination of autosegmentation of the proximal femur and machine learning analysis with a reference standard of dual-energy X-ray absorptiometry (DXA).

**Methods:** Abdomen-pelvic CT scans were retrospectively analyzed from 1,122 patients who received both DXA and abdomen-pelvic computed tomography (APCT) scan from January 2018 to December 2020. The study cohort consisted of a training cohort and a temporal validation cohort. The left proximal femur was automatically segmented, and a prediction model was built by machine-learning analysis using a random forest (RF) analysis and 854 PyRadiomics features. The technical success rate of autosegmentation, diagnostic test, area under the receiver operator characteristics curve (AUC), and precision recall curve (AUC-PR) analysis were used to analyze the training and validation cohorts.

**Results:** The osteoporosis prevalence of the training and validation cohorts was 24.5%, and 10.3%, respectively. The technical success rate of autosegmentation of the proximal femur was 99.7%. In the diagnostic test, the training and validation cohorts showed 78.4% *vs.* 63.3% sensitivity, 89.4% *vs.* 98.1% specificity. The prediction performance to identify osteoporosis within the groups used for training and validation cohort was high and the AUC and AUC-PR to forecast the occurrence of osteoporosis within the training and validation cohorts were 90.8% [95% confidence interval (CI), 88.4–93.2%] *vs.* 78.0% (95% CI, 76.0–79.9%) and 94.6% (95% CI, 89.3–99.8%) *vs.* 88.8% (95% CI, 86.2–91.5%), respectively.

**Conclusions:** The osteoporosis prediction model using autosegmentation of proximal femur and machine-learning analysis with PyRadiomics features on APCT showed excellent diagnostic feasibility and technical success.

**Keywords:** Diagnosis; computer-assisted; osteoporosis; tomography; X-ray computed; machine-learning

Submitted Dec 11, 2023. Accepted for publication Apr 07, 2024. Published online May 08, 2024.

doi: 10.21037/qims-23-1751

View this article at: <https://dx.doi.org/10.21037/qims-23-1751>

<sup>^</sup> ORCID: Min Su Park, 0000-0002-3356-203X; Hong Il Ha, 0000-0001-9050-7486; Hyun Kyung Lim, 0000-0003-2319-3135; Junhee Han, 0000-0001-5736-8752; Seongyong Pak, 0000-0002-2978-2355.

## Introduction

As the life expectancy increases, osteoporosis has become a major public health concern worldwide (1-3). Osteoporosis is a very common disease in middle-aged and older women, and 40% of women over 70 years of age are diagnosed with osteoporosis. Half of osteoporotic people will have at least one major osteoporosis-related fracture during the rest of their lifetime (1,3,4). Proper management through an early diagnosis of osteoporosis can have a significant impact on patient prognosis (5). Nevertheless, osteoporosis is a silent disease, and people frequently fail to acknowledge the gravity of this illness until a major fracture occurs (6,7). As a result, asymptomatic people do not participate in the screening program, and this results in the underuse of dual-energy X-ray absorptiometry (DXA) for osteoporosis (8-12).

There have been growing efforts and attempt to improve the screening of osteoporosis and to overcome the limitations and underuse of DXA (13-17). In the era of artificial intelligence-related research and radiomics analysis, the prediction of osteoporosis or the use of the bone mineral density (BMD) with abdomen-pelvic computed tomography (APCT) has been a new paradigm in opportunistic osteoporosis screening research (18-21). Despite the high prediction performance of these values which was suggested as maximal area under the receiver operator characteristics curve (AUC) of 0.74 by Buckens *et al.*, the main weakness of previous studies was that the area or volume of interest must be manually drawn or segmented to analyze these values (15,16,20,22). Advances in technology have made it possible to automatically segment the proximal femur volume and to analyze 854 PyRadiomics features simultaneously. Radiomics feature analysis may be a suitable technique to evaluate microstructural changes in trabecular bone, including the density, shape, size, and interactions of imaging features with or without wavelet transformation (18,19,22-26). In addition, artificial intelligence in medical practice has been proven effective in big data-based screening (27-30). Thus, the aim of this study was to assess the diagnostic performance of an osteoporosis prediction model made by a combination of autosegmentation of the proximal femur and machine learning analysis with 854 PyRadiomics features from precontrast APCT with a reference standard of DXA. We present this article in accordance with the TRIPOD reporting checklist (31) (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1751/rc>).

## Methods

This retrospective study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional review board of Hallym University Sacred Heart Hospital (No. HALLYM 2020-12-015), and the need for informed consent was waived due to the nature of the retrospective analysis.

### Patients

#### Training and validation cohort patients

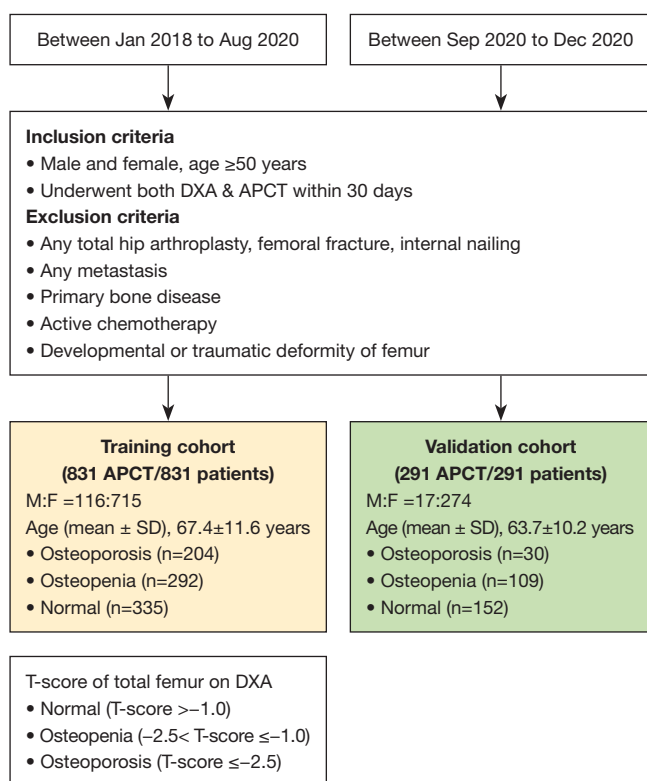
From January 2018 to August 2020, 915 individuals aged 50 years or older underwent both APCT and DXA scans within a 1-month period (mean,  $3.8 \pm 6.1$  days; range, 0–30 days) at Hallym University Sacred Heart Hospital. Among them, 84 patients were excluded due to various reasons: bone metastases ( $n=11$ ), metastasis outside the bone ( $n=13$ ), recent chemotherapy within the past three months prior to the computed tomography (CT) scan ( $n=28$ ), primary bone diseases (e.g., fibrous dysplasia;  $n=5$ ), developmental or traumatic femoral deformities ( $n=8$ ), previous total hip arthroplasty, femoral fractures, or internal nailing ( $n=16$ ), or incomplete scans ( $n=3$ ). Between September and December 2020, 291 consecutive patients were included as the temporal validation cohort, following the same enrollment criteria as the training cohort (*Figure 1*).

### DXA

A solitary scanner was used for the DXA procedure (GE Healthcare Lunar Prodigy Densitometers, Germany). The T-score of the total femur among seven values (neck, upper neck, lower neck, Ward's triangle, trochanter, shaft, and total) was used as the reference standards because the segmentation volume of the femur was based on the total femur area on the DXA. The total femur T-score was interpreted as either osteoporosis (T-score  $\leq -2.5$ ), osteopenia ( $-2.5 < \text{T-score} < -1.0$ ), or normal (T-score  $\geq -1.0$ ) (32).

### CT imaging

We conducted CT examinations using three multidetector-row CT scanners (SOMATOM Definition Edge, SOMATOM Definition Flash, SOMATOM Force; Siemens Healthineers, Germany) operating in standard single-energy CT mode. The pixel sizes were approximately 0.634 mm,



**Figure 1** Patient enrollment flowchart. DXA, dual-energy X-ray absorptiometry; APCT, abdomen-pelvic computed tomography; M, male; F, female; SD, standard deviation.

with voxel sizes around  $0.317 \text{ mm}^3$ . While minor variations may exist between patients, these differences had a negligible impact on the study. We utilized Automatic tube voltage selection (Care kVp) and automatic tube current modulation (CARE Dose 4D) protocols. To minimize the influence of contrast agent administration, all measurements were performed using CT images acquired before contrast administration. The scanning parameters included detector collimations of  $128 \times 0.6$  or  $192 \times 0.55 \text{ mm}$ , pitch ranging from 0.6 to 0.9, gantry rotation time of 0.5 s, tube current between 75–659 mAs, tube voltage of 100–120 kVp, and iterative reconstruction.

### Radiomics analysis

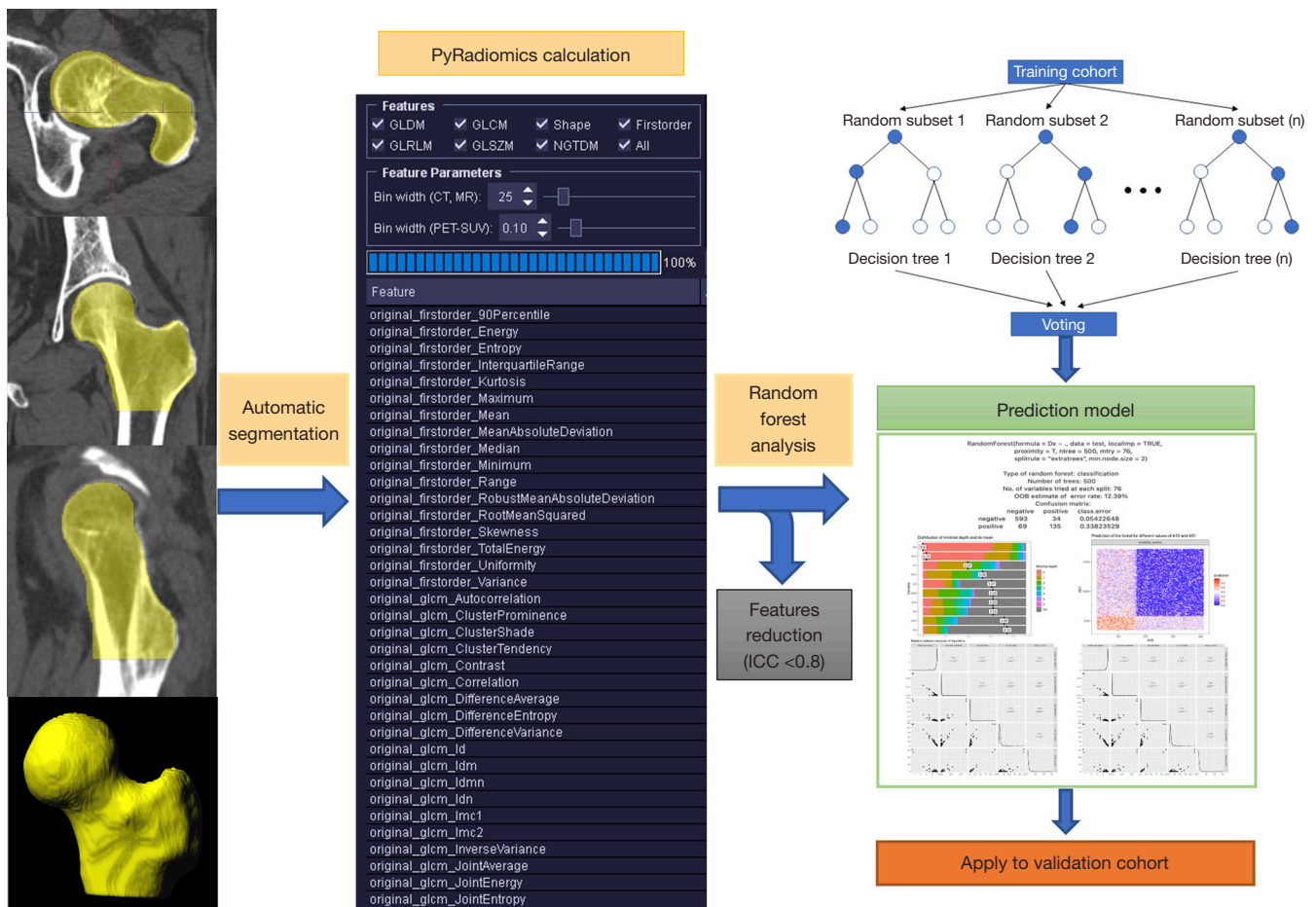
Proximal femur segmentation was performed by one radiologist with 15 years of experience in radiology using a dedicated radiomics analysis software, which automatically segmented the proximal femur from the femoral head to the lower trochanter level (above the horizontal line of

the ischial tuberosity lower margin) (syngo.via Frontier Radiomics, version 1.2.6, Siemens Healthineers) (Figure 2). Deep learning based autosegmentation, which took less than a minute was conducted to align with the reference test, DXA, and scan range, thereby reducing potential overestimation caused by individual anatomical variations. Furthermore, based on the previous study, Hounsfield unit histogram analysis (HUHA) value at the femur neck and osteoporosis demonstrated a strong correlation (16).

The technical failure rate was calculated to evaluate the accuracy of the autosegmentation. A technical failure was defined as a case in which the target femur volume was not extracted for more than 90% of the original volume or when areas other than the femur region was extracted. A total of 854 PyRadiomics features were based on PyRadiomics (<https://pyradiomics.readthedocs.io/>) (33). All features consisted of the following four classes: (I) 17 shape and size features; (II) 18 first-order statistics, features of the distribution of voxel intensities; (III) 24 gray-level cooccurrence matrix features; (IV) 14 gray-level dependence matrix features; (V) 16 gray-level run-length matrix features; (VI) 16 gray-level size zone matrix features; (VII) 17 neighboring gray tone difference matrix features; and (VIII) 744 wavelet transformation features of first-order statistics and texture features. The wavelet transformation efficiently separates the textural details by breaking down the original image, akin to how Fourier analysis operates at different frequencies—both low and high.

### Building the osteoporosis prediction model using a RF analysis

We employed the random undersampling algorithm, a widely used data-processing technique in machine learning, on the training cohort to mitigate the imbalance between osteoporosis statuses (osteoporosis *vs.* non-osteoporosis, 234 APCT scans/889 APCT scans). This approach aimed to prevent bias towards cases in the majority class and achieve higher classification accuracy. The RF method was chosen for building the prediction model due to its effective variance-bias trade-off. To address repeatability issues in radiomics, features with an intraclass correlation coefficient (ICC) above 0.8 were deemed stable and selected for model construction. Further feature selection was omitted to reduce computation time, prioritizing efficiency over incremental prediction accuracy, particularly with the RF method. We



**Figure 2** Workflow of the osteoporosis prediction model built with the automatic segmentation and machine learning analysis using PyRadiomics features. The target femur was automatically segmented. 854 PyRadiomics features, such as the voxel intensities, shape, size, and texture features, were calculated and analyzed after cross checking the segmentation volume. After the reduction of the PyRadiomics features by an intraclass correlation coefficient greater than 0.8, the prediction model was built by a random forest machine learning analysis. ICC, intraclass correlation coefficient.

employed a fivefold cross-validation to enhance model efficacy, optimizing hyperparameters such as the number of trees in the forest and minimum sample size for leaf nodes through cross-validation.

### Statistical analysis

We conducted all statistical analyses using R software (version 3.6.1; R Foundation for Statistical Computing: <http://www.Rproject.org>) and MedCalc for Windows, version 20.015 (MedCalc Software). Primarily, we employed two-sided tests. The reproducibility of segmented volume and radiomics features was assessed using ICCs with a two-

way random model for absolute measurements. To evaluate reproducibility, a radiologist with 13 years of experience measured 50 randomly selected CT scans. We calculated prediction accuracy for both the training and validation cohorts using the diagnostic test confusion matrix. The performance of the radiomics prediction model in both cohorts was assessed using AUC and precision recall curve (AUC-PR). AUCs between cohorts were compared using the DeLong method. AUC-PR determines whether the prediction model can accurately identify all positive examples without falsely classifying too many negative examples as positive. A significance level of  $P < 0.05$  was considered statistically significant.

**Table 1** Comparison of the patient demographics of the training and validation cohorts

Index	Training cohort (831 APCT)	Validation cohort (291 APCT)	P value
Overall (n=1122)			
Sex (M:F)	116:715	17:274	<0.001
Age (years)	67.4±11.6	63.7±10.2	<0.001
BMI (kg/m <sup>2</sup> )	23.9±3.9	25.0±9.5	0.006
Days btw DXA & APCT	3.6±6.0	3.9±5.4	0.78
T-score (overall)	-1.4±1.3	-1.0±1.3	<0.001
BMD (g/cm <sup>2</sup> )	0.764±0.161	0.850±0.141	<0.001
Osteoporosis (n=234)	204	30	
T-score	-3.1±0.5	-3.5±1.9	0.06
BMD (g/cm <sup>2</sup> )	0.560±0.069	0.560±0.087	0.009
Osteopenia (n=401)	292	109	
T-score	-1.7±0.4	-1.7±0.3	0.06
BMD (g/cm <sup>2</sup> )	0.728±0.051	0.771±0.060	<0.001
Normal (n=487)	335	152	
T-score	-0.2±0.7	-0.20.7	0.98
BMD (g/cm <sup>2</sup> )	0.919±0.090	0.954±0.083	<0.001

All values are presented as mean ± standard deviation, or number (frequency). APCT, abdomen-pelvic computed tomography; M, male; F, female; BMD, bone material density; BMI, body mass index; DXA, dual-energy X-ray absorptiometry.

## Results

### Demographics of the study population

The demographics of the training and validation cohorts are summarized in *Table 1*. Overall, 204 patients were diagnosed with osteoporosis in the training cohort, and 30 patients were diagnosed with osteoporosis in the validation cohort. The osteoporosis prevalence of each cohort was 24.5% and 10.3%, respectively. Most demographic variables were significantly different, but the T-score as a standard reference showed no difference in each of the subgroup in both cohorts, as shown in *Table 1* ( $P>0.05$ ).

### The technical failure rate and reproducibility of the autosegmentation

Of the 1,125 APCT scans, three APCT scans failed the autosegmentation. Thus, the technical failure rate was

approximately 0.3%. The autosegmentation failed in patients whose ischial tuberosity was not identified due to hip flexion, patients being in an irregular position rather than the supine position, or the presence of severe thoracolumbar scoliosis.

The proximal femur volume, as determined by autosegmentation, showed almost perfect agreements [ICC =0.99, 95% confidence interval (CI), 0.99–0.99]. A total of 669 wavelet transformation features were excluded because of having ICC less than 0.1. A total of 185 features (110 original first-order features and 75 wavelet transformation features) were selected and showed almost perfect reproducibility (ICC =0.96, 95% CI, 0.94–0.97).

### Important radiomics features in the machine learning prediction model

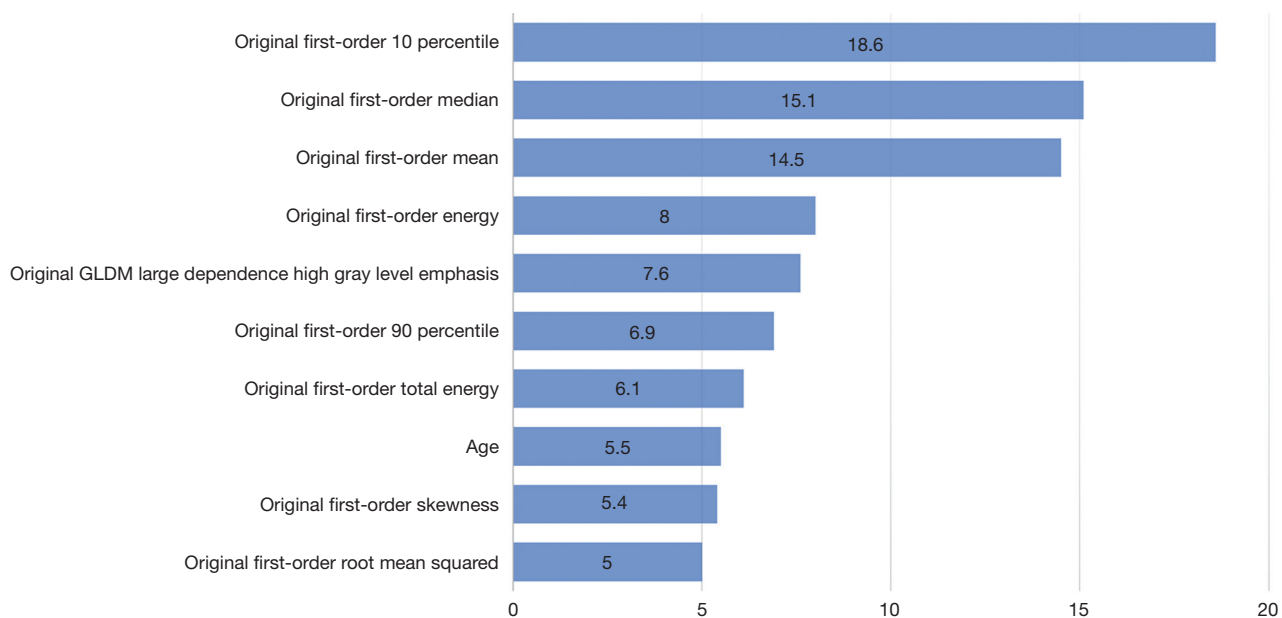
Of the 185 PyRadiomics features with age and sex included in the RF prediction model, the top 10 important features are displayed by the Mean Decrease Gini (*Figure 3*).

### Diagnostic performance of the osteoporosis prediction model

The prediction performance of the model according to each subcategory group is summarized in *Table 2*. Approximately 1% of the normal cases in both cohorts were predicted to be osteoporosis cases. Overall, the percentage of false positive cases were 4.5% (37/831) and 1.7% (5/291) in the training and validation cohorts, respectively; 7 of 487 normal cases and 35 of 401 osteopenia cases in the respective cohorts were mispredicted as osteoporosis. The overall diagnostic accuracy of the prediction model in the training cohort and its 5-fold cross validations as well as in the validation cohort are summarized in *Table 3*. In the diagnostic test, the training and validation cohorts showed 78.4% *vs.* 63.3% sensitivity, 89.4% *vs.* 98.1% specificity, 94.1% *vs.* 95.9% NPV, 65.7% *vs.* 79.2% PPV, and 87.1% *vs.* 94.5% accuracy, respectively. The AUCs of the training and validation cohorts to predict femoral osteoporosis were 90.8% (95% CI, 88.4–93.2%) and 94.6% (95% CI, 89.3–99.8%), respectively, without a significant difference ( $P=0.20$ ) (*Figure 4*). The AUC-PR of the training and validation cohorts was 78.0% (95% CI, 76.0–79.9%) and 88.8% (95% CI, 86.2–91.5%), respectively (*Figure 5*).

## Discussion

The main aim of this study was to assess the prediction



**Figure 3** Top 10 important radiomics features by the Mean Decrease Gini. The unit for age is years. GLDM, gray level dependence matrix.

**Table 2** Prediction performance according to each subcategory of 1,122 APCT cases

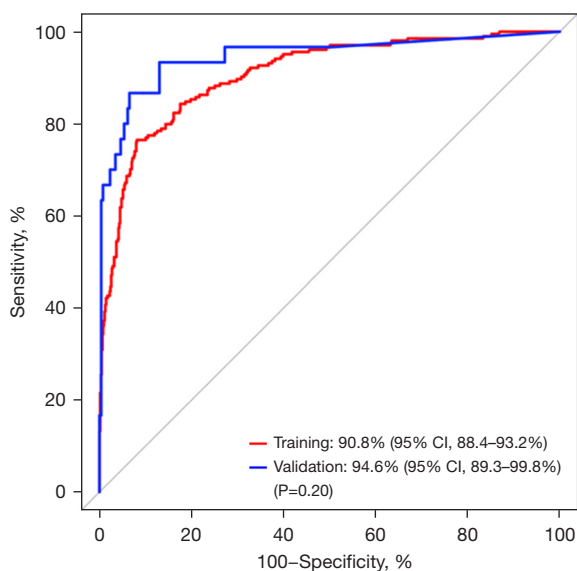
Diagnosis	Prediction (Y = osteoporosis, N = non-osteoporosis)					
	Normal (n=487)		Osteopenia (n=401)		Osteoporosis (n=234)	
	Yes	No	Yes	No	Yes	No
Osteoporosis	0 (0)	0 (0)	0 (0)	0 (0)	134 (19)	70 (11)
Non-osteoporosis	4 (3)	330 (136)	33 (2)	260 (120)	0 (0)	0 (0)

Each number in parenthesis is the data of the validation cohort. APCT, abdomen-pelvic computed tomography.

**Table 3** The diagnostic accuracy of the prediction model of the 5-fold cross validation of the training cohort, overall training cohort and validation cohort

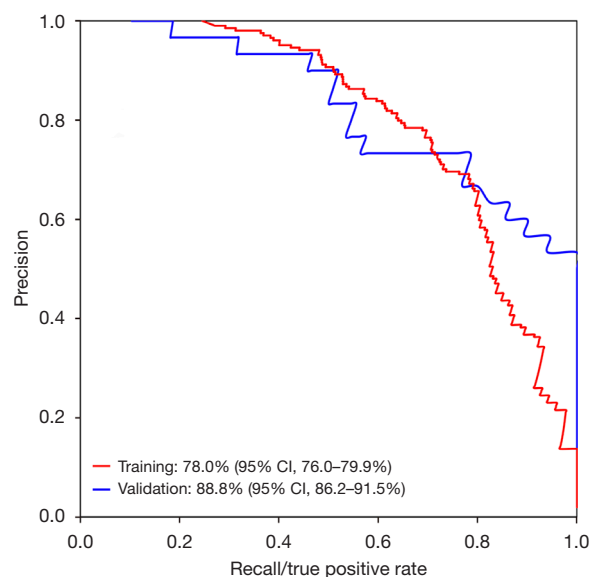
Dataset	True positive (n)	True negative (n)	False positive (n)	False negative (n)	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Accuracy (%) (95% CI)
CV_1st	23	119	6	18	56.1	95.2	86.9	79.3	85.5 (79.3–90.5)
CV_2nd	31	117	9	10	75.6	92.9	92.1	77.5	88.6 (82.8–93.0)
CV_3rd	23	122	4	18	56.1	96.8	87.1	85.2	86.8 (80.7–91.6)
CV_4th	29	116	9	12	70.7	92.8	90.6	76.3	87.4 (81.3–92.0)
CV_5th	25	114	11	15	66.5	91.2	88.4	69.4	84.2 (77.8–89.4)
Training cohort	134	590	70	37	78.4	89.4	94.1	65.7	87.1 (84.7–89.3)
Validation cohort	19	256	5	11	63.3	98.1	95.9	79.2	94.5 (91.2–96.8)

NPV, negative predictive value; PPV, positive predictive value; CI, confidence interval; CV, cross validation.



**Figure 4** Comparison of the area under the curves of the training cohort (AUC =90.8%) and validation cohort (AUC =94.6%) (P=0.20). CI, confidence interval; AUC, area under the receiver operator characteristics curve.

of femoral osteoporosis by employing autosegmentation and machine-learning analysis with PyRadiomics features based on APCT scans. In this study, autosegmentation of the proximal femur showed a 99.7% technical success rate. In addition, the prediction performance of the RF model (AUC) was 90.8% and 94.6% in the training and validation cohorts, respectively, which was uncommon but probably due to small sample size of the study. The high specificity and high NPV were considered meaningful results to select healthy people and to reduce unnecessary DXA tests. The precision-recall curve illustrates the balance between precision and recall across various thresholds. A substantial area under the curve signifies both heightened recall and precision. Elevated precision corresponds to a diminished false positive rate, while heightened recall aligns with a reduced false negative rate. In particular, since the AUC-PR is not affected by the number of true negative patients, it is known that the AUC-PR is better than the AUC in imbalanced populations (34). In this study, the prevalence of osteoporosis was significantly different between the training and validation cohorts. However, the AUC-PR and AUC of the validation cohort were superior to those of the training cohorts, and therefore, this prediction model proved the feasibility of the diagnostic performance of the model in



**Figure 5** The area under the precision-recall curves of the prediction model in the training cohort and validation cohort. The 95% confidence intervals are rounded values. CI, confidence interval.

both the AUC and AUC-PR analyses.

Most of the important radiomics features overlapped with a prior study (20), except for the negative HU value that was assumed for the fatty marrow. An increased fatty marrow content is known as the most important pathophysiologic change in osteoporosis, and the important features have been analyzed in previous studies (16,20,35). Although the PyRadiomics features includes the 10 percentiles of HU, this feature would not be sufficient to accurately reflect the fatty bone marrow changes in each case. In addition, we examined the volume of the femur extending from the head to the lesser trochanter, as this region aligns with the total femoral area in DXA scans. In this study, the T-score of the total femoral volume was used as a reference standard instead of the lowest T-score of femur, which has been applied in the previous studies (16,20). These differences may be reasons for the relatively low sensitivity of this study. However, considering the high diagnostic performance and near perfect technical success rate of autosegmentation, osteoporosis prediction using this study model could be applied in clinical practice.

This prediction model was designed for a binary classification of cases with osteoporosis or non-osteoporosis, which consists of normal cases and osteopenia cases.

Therefore, there may be concerns that normal cases were incorrectly predicted as osteoporosis. We hypothesized that a model capable of accurately predicting osteoporosis would have no problem in distinguishing normal femurs. Considering that the T-score varies according to age and sex even if the BMD value is the same, it is difficult and challenging to distinguish osteoporosis from non-osteoporosis cases (36). In the outcome analysis of the training and validation cohorts, approximately 1–2% of normal cases were mispredicted as osteoporosis with the machine learning model made by only the PyRadiomics features. The precise tuning of important variables, such as the target femur volume, PyRadiomics features, and HU histogram analysis, in addition to the adjustment for age and sex, could improve the prediction model performance.

Prior studies have shown the usefulness of the HU histogram analysis alone and combination of using the HU histogram analysis and radiomics features (16,20). Although these studies showed feasibility to predict osteoporosis with high sensitivity and specificity and with an accuracy up to 95%, the segmentation of the proximal femur and radiomics analysis in addition to the HU histogram were performed manually. It is a time-consuming task to apply this model to real-time clinical practice. However, in this study, all segmentation and analysis of the target volume were automatically performed within a few minutes. This is not only a sufficient time for clinical application without additional workload but can also remove the measurement error-related observer bias.

Even though there was almost perfect agreement of the autosegmented volume on the ICC analysis, 669 PyRadiomics feature-related wavelet transformations were excluded because of low reproducibility. In prior study, the wavelet transformation features showed poor reproducibility on the different radiomics analysis software programs (20). Thus, the wavelet transformation features were considered insignificant features in the femoral osteoporosis analysis.

Age and female sex have been considered important variables in osteoporosis (4,37). Interestingly, these factors were ranked lower on the feature importance evaluation on the machine learning analysis. The PyRadiomics features were affected and were linked according to the bony microstructure change depending on the BMD. These changes were dependent on age and sex, so they were already reflected in the important PyRadiomics features. Thus, age and sex seem to be ranked at a lower level.

The limitation of this study was that there was an

imbalance in the sex ratio of osteoporosis patients and that this was a single-center study. Osteoporosis is a consequence of aging and rapidly progresses after 50 years of age in women, so a gender imbalance is inevitable (35). To handle these imbalances, a random undersampling algorithm was applied to build a prediction model with a RF analysis, and an AUC-PR analysis was added. The PyRadiomics features are significantly affected by the CT acquisition technique, and these included the tube voltage, tube current, reconstruction algorithm and manufacturer. This heterogeneity of CT acquisition status is a large barrier for the external validation. Therefore, in this study, the prediction model had to be verified using the temporal validation cohort, but a prospective multicenter study is being considered by referring to the results of this study and the existing HU histogram analysis studies (16,20). Based solely on the research findings, it is expected that the results would apply well to excluded patients. However, there may be conditions among excluded patients such as metastasis or primary bone diseases that could introduce confusion into the calculations. Therefore, this aspect may need to be addressed in future studies. Lastly, fracture risk indicators that cannot be assessed by areal BMD or CT-HU alone are currently the focus of attention in the osteoporosis field and these areas are important for future research endeavors, which we plan to explore in subsequent studies (38).

## Conclusions

In conclusion, femoral osteoporosis prediction by the combination of autosegmentation and machine learning analysis using PyRadiomics and APCT proved to be an opportunistic screening feasibility with more than 90% accuracy, specificity, and NPV as well as a 99.7% of technical success rate of autosegmentation.

## Acknowledgments

*Funding:* This work was supported by the Central Medical Service (CMS) Research Fund. The specific grant number was not assigned by the company or funder (Central Medical Service Company, Ltd., Seoul, Korea).

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://qims>.



[amegroups.com/article/view/10.21037/qims-23-1751/rc](https://amegroups.com/article/view/10.21037/qims-23-1751/rc)

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1751/coif>). All authors report that this work was supported by the Central Medical Service (CMS) Research Fund. S.P. is an employee of Siemens-Healthineers throughout his involvement in the study. He supported the dedicated analysis program known as syngo.via Frontier Radiomics, version 1.2.6, Siemens Healthineers. As an employee of Siemens and a research collaborator, he provided support in installing and using a dedicated analysis program during the research period. He did not participate in the analysis of the study subjects and results. The results of this study are not discussed or shared with Siemens-Healthineers. The authors have no other conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This retrospective study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional review board of Hallym University Sacred Heart Hospital (No. HALLYM 2020-12-015), and the need for informed consent was waived due to the nature of the retrospective analysis.

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

## References

- Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. *J Bone Miner Res* 2007;22:465-75.
- Dempster DW. Osteoporosis and the burden of osteoporosis-related fractures. *Am J Manag Care* 2011;17 Suppl 6:S164-9.
- Sambrook P, Cooper C. Osteoporosis. *Lancet* 2006;367:2010-8.
- Curtis EM, Moon RJ, Harvey NC, Cooper C. The impact of fragility fracture and approaches to osteoporosis risk assessment worldwide. *Bone* 2017;104:29-38.
- Osteoporosis prevention, diagnosis, and therapy. *JAMA* 2001;285:785-95.
- Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, Cauley J, Black D, Vogt TM. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med* 1995;332:767-73.
- Kern LM, Powe NR, Levine MA, Fitzpatrick AL, Harris TB, Robbins J, Fried LP. Association between screening for osteoporosis and the incidence of hip fracture. *Ann Intern Med* 2005;142:173-81.
- Crandall CJ, Larson J, Gourlay ML, Donaldson MG, LaCroix A, Cauley JA, Wactawski-Wende J, Gass ML, Robbins JA, Watts NB, Ensrud KE. Osteoporosis screening in postmenopausal women 50 to 64 years old: comparison of US Preventive Services Task Force strategy and two traditional strategies in the Women's Health Initiative. *J Bone Miner Res* 2014;29:1661-6.
- Nelson HD, Haney EM, Dana T, Bougatsos C, Chou R. Screening for osteoporosis: an update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2010;153:99-111.
- Riggs BL, Melton LJ 3rd. The worldwide problem of osteoporosis: insights afforded by epidemiology. *Bone* 1995;17:505S-11S.
- Raisz LG. Clinical practice. Screening for osteoporosis. *N Engl J Med* 2005;353:164-71.
- Jiang YW, Xu XJ, Wang R, Chen CM. Radiomics analysis based on lumbar spine CT to detect osteoporosis. *Eur Radiol* 2022;32:8019-26.
- Tay WL, Chui CK, Ong SH, Ng AC. Osteoporosis screening using areal bone mineral density estimation from diagnostic CT images. *Acad Radiol* 2012;19:1273-82.
- 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. *Ann Intern Med* 2013;158:588-95.
- Buckens CF, Dijkhuis G, de Keizer B, Verhaar HJ, de Jong PA. Opportunistic screening for osteoporosis on routine computed tomography? An external validation study. *Eur Radiol* 2015;25:2074-9.

16. Lim HK, Ha HI, Park SY, Lee K. Comparison of the diagnostic performance of CT Hounsfield unit histogram analysis and dual-energy X-ray absorptiometry in predicting osteoporosis of the femur. *Eur Radiol* 2019;29:1831-40.
17. Ziemlewicz TJ, Binkley N, Pickhardt PJ. Opportunistic Osteoporosis Screening: Addition of Quantitative CT Bone Mineral Density Evaluation to CT Colonography. *J Am Coll Radiol* 2015;12:1036-41.
18. van Hamersvelt RW, Schilham AMR, Engelke K, den Harder AM, de Keizer B, Verhaar HJ, Leiner T, de Jong PA, Willeminck MJ. Accuracy of bone mineral density quantification using dual-layer spectral detector CT: a phantom study. *Eur Radiol* 2017;27:4351-9.
19. Rastegar S, Vaziri M, Qasempour Y, Akhash MR, Abdalvand N, Shiri I, Abdollahi H, Zaidi H. Radiomics for classification of bone mineral loss: A machine learning study. *Diagn Interv Imaging* 2020;101:599-610.
20. Lim HK, Ha HI, Park SY, Han J. Prediction of femoral osteoporosis using machine-learning analysis with radiomics features and abdomen-pelvic CT: A retrospective single center preliminary study. *PLoS One* 2021;16:e0247330.
21. Jang S, Graffy PM, Ziemlewicz TJ, Lee SJ, Summers RM, Pickhardt PJ. Opportunistic Osteoporosis Screening at Routine Abdominal and Thoracic CT: Normative L1 Trabecular Attenuation Values in More than 20 000 Adults. *Radiology* 2019;291:360-7.
22. Alacreu E, Moratal D, Arana E. Opportunistic screening for osteoporosis by routine CT in Southern Europe. *Osteoporos Int* 2017;28:983-90.
23. Baum T, Grande Garcia E, Burgkart R, Gordijenko O, Liebl H, Jungmann PM, Gruber M, Zahel T, Rummeny EJ, Waldt S, Bauer JS. Osteoporosis imaging: effects of bone preservation on MDCT-based trabecular bone microstructure parameters and finite element models. *BMC Med Imaging* 2015;15:22.
24. Bauer JS, Sidorenko I, Mueller D, Baum T, Issever AS, Eckstein F, Rummeny EJ, Link TM, Raeth CW. Prediction of bone strength by  $\mu$ CT and MDCT-based finite-element-models: how much spatial resolution is needed? *Eur J Radiol* 2014;83:e36-42.
25. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* 2020;20:33.
26. Pfähler E, Zhovannik I, Wei L, Boellaard R, Dekker A, Monshouwer R, El Naqa I, Bussink J, Gillies R, Wee L, Traverso A. A systematic review and quality of reporting checklist for repeatability and reproducibility of radiomic features. *Phys Imaging Radiat Oncol* 2021;20:69-75.
27. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016;375:1216-9.
28. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563-77.
29. Peng T, Zeng X, Li Y, Li M, Pu B, Zhi B, Wang Y, Qu H. A study on whether deep learning models based on CT images for bone density classification and prediction can be used for opportunistic osteoporosis screening. *Osteoporos Int* 2024;35:117-28.
30. Pickhardt PJ, Correale L, Hassan C. AI-based opportunistic CT screening of incidental cardiovascular disease, osteoporosis, and sarcopenia: cost-effectiveness analysis. *Abdom Radiol (NY)* 2023;48:1181-98.
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;13:1.
32. Lewiecki EM, Gordon CM, Baim S, Leonard MB, Bishop NJ, Bianchi ML, Kalkwarf HJ, Langman CB, Plotkin H, Rauch F, Zemel BS, Binkley N, Bilezikian JP, Kendler DL, Hans DB, Silverman S. International Society for Clinical Densitometry 2007 Adult and Pediatric Official Positions. *Bone* 2008;43:1115-21.
33. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017;77:e104-7.
34. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.
35. Poole KES, Skingle L, Gee AH, Turmezei TD, Johannesdottir F, Blesic K, Rose C, Vindlacheruvu M, Donell S, Vaculik J, Dungal P, Horak M, Stepan JJ, Reeve J, Treece GM. Focal osteoporosis defects play a key role in hip fracture. *Bone* 2017;94:124-34.
36. Blake GM, Fogelman I. The clinical role of dual energy X-ray absorptiometry. *Eur J Radiol* 2009;71:406-14.
37. Carpenter RD, Sigurdsson S, Zhao S, Lu Y, Eiriksdottir G, Sigurdsson G, Jonsson BY, Prevrhal S, Harris TB, Siggeirsdottir K, Guðnason V, Lang TF. Effects of age and

- sex on the strength and cortical thickness of the femoral neck. *Bone* 2011;48:741-7.
38. Aggarwal V, Maslen C, Abel RL, Bhattacharya P, Bromiley PA, Clark EM, Compston JE, Crabtree N, Gregory JS, Kariki EP, Harvey NC, Ward KA, Poole

KES. Opportunistic diagnosis of osteoporosis, fragile bone strength and vertebral fractures from routine CT scans; a review of approved technology systems and pathways to implementation. *Ther Adv Musculoskelet Dis* 2021;13:1759720X211024029.

**Cite this article as:** Park MS, Ha HI, Lim HK, Han J, Pak S. Femoral osteoporosis prediction model using autosegmentation and machine learning analysis with PyRadiomics on abdomen-pelvic computed tomography (CT). *Quant Imaging Med Surg* 2024;14(6):3959-3969. doi: 10.21037/qims-23-1751