Published in final edited form as:
J Electron Imaging. 2014 February 4; 23(1): 013013–. doi:10.1117/1.JEI.23.1.013013.

Improving bone strength prediction in human proximal femur specimens through geometrical characterization of trabecular bone microarchitecture and support vector regression

Chien-Chun Yang[#a,*], **Mahesh B. Nagarajan**[#a], **Markus B. Huber**[a], **Julio Carballido-Gamio**[b], **Jan S. Bauer**[c], **Thomas Baum**[c], **Felix Eckstein**[d], **Eva Lochmüller**[d], **Sharmila Majumdar**[b], **Thomas M. Link**[b], and **Axel Wismüller**[a,e]

[a]University of Rochester, Departments of Imaging Sciences and Biomedical Engineering, Rochester, New York 14627

[b]University of California San Francisco, Musculoskeletal and Quantitative Imaging Research, Department of Radiology and Biomedical Imaging, San Francisco, California 94143

[c]Technische Universität München, Institut Für Röntgendiagnostik, Munich, München 85748, Germany

[d]Paracelsus Medical University Salzburg, Institute of Anatomy and Musculoskeletal Research, Salzburg 5020, Austria

[e]University of Munich, Department of Radiology, München 80539, Germany

[#] These authors contributed equally to this work.

Abstract

We investigate the use of different trabecular bone descriptors and advanced machine learning tech niques to complement standard bone mineral density (BMD) measures derived from dual-energy x-ray absorptiometry (DXA) for improving clinical assessment of osteoporotic fracture risk. For this purpose, volumes of interest were extracted from the head, neck, and trochanter of 146 *ex vivo* proximal femur specimens on multidetector computer tomography. The trabecular bone captured was characterized with (1) statistical moments of the BMD distribution, (2) geometrical features derived from the scaling index method (SIM), and (3) morphometric parameters, such as bone fraction, trabecular thickness, etc. Feature sets comprising DXA BMD and such supplemental features were used to predict the failure load (FL) of the specimens, previously determined through biomechanical testing, with multiregression and support vector regression. Prediction performance was measured by the root mean square error (RMSE); correlation with measured FL was evaluated using the coefficient of determination R^2. The best prediction performance was achieved by a combination of DXA BMD and SIM-derived geometric features derived from the femoral head (RMSE: 0.869 ± 0.121, R^2: 0.68 ± 0.079), which was significantly better than DXA BMD alone (RMSE: 0.948 ± 0.119, R^2: 0.61 ± 0.101) ($p < 10^{-4}$). For multivariate feature sets, SVR outperformed multiregression ($p < 0.05$). These results suggest

[*]Address all correspondence to: Chien-Chun Yang, chien-chun.yang@rochester.edu.

that supplementing standard DXA BMD measurements with sophisticated femoral trabecular bone characterization and supervised learning techniques can significantly improve biomechanical strength prediction in proximal femur specimens.

Keywords

osteoporosis; trabecular bone; dual x-ray absorptiometry; bone mineral density; quantitative computer tomography; scaling index method; support vector regression

1 Introduction

Osteoporosis is one of the most common age-related diseases among elderly people. The progression of osteoporosis can lead to osteoporotic fractures, which reduces the quality of life and increases the mortality rate. Previous studies have predicted that the number of people at risk for osteoporotic fracture will reach 6.26 million worldwide by the year 2050.[1,2] Thus, accurate prediction of osteoporotic fracture risks can be an important aid for clinical assessment and management of osteoporosis.

Dual-energy x-ray absorptiometry (DXA) has been used as a standard technique for measuring bone quality in terms of bone mineral density (BMD) for purposes of osteoporotic fracture risk estimation.[3–6] BMD measurements through DXA at the site of the proximal femur have shown to be highly predictive of bone fractures when compared to other sites.[3–5,7] However, DXA-derived BMD measurements are affected by interference from surrounding cortical shell, adipose tissue, and soft tissue, which can result in inaccuracies in bone strength estimation.[6–11] Quantitative computer tomography (QCT) can overcome such shortcomings by eliminating interference from the surrounding tissue and provide an exclusive measure of BMD in the trabecular bone compartment. In fact, the ability of QCT to improve the efficacy of fracture risk assessment, by excluding sources of error, such as osteophytes and hypertrophic posterior elements, which may artificially elevate integral BMD measures, has been previously demonstrated in spinal fracture studies.[11–14]

While QCT-based BMD measures exhibit a strong correlation with fracture risk, they still do not serve as reliable predictors of bone strength. Previous work has shown that QCT-based BMD measurements for patients with and without prevalent femur fractures overlap,[15,16] which is further supported by observations that QCT-based BMD measurements within the average normal range only suggests a lower probability of developing osteoporosis or related fractures.[11–14,17] This inconsistency could stem from the fact that such QCT-based BMD measures do not account for a complete profile of morphological and structural variations in trabecular bone microarchitecture. In this regard, other studies have investigated the inclusion of anatomical variables such as bone volume, femoral head axis length, and other structural parameters in addition to BMD.[18–20] More recently, textural approaches involving gray-level co-occurrence matrices or geometrical features derived from the scaling index method (SIM) have been proposed to characterize trabecular bone microarchitecture and complement conventional DXA and QCT-based BMD measures.[18,19,21–23]

In this study, we propose, implement, and evaluate an automated approach to bone strength prediction that includes both conventional DXA BMD measurements and a complete characterization of trabecular bone microarchitecture on multidetector CT (MDCT) images. Furthermore, such characterization extracted from a training subset of *ex vivo* proximal femur specimens is used to construct bone strength prediction models with advanced machine learning techniques. We specifically focus on the use of support vector regression (SVR) over traditionally used multiregression. These models are subsequently evaluated on an independent test set of femur specimens for their ability to predict bone strength. This application of supervised learning allows us to evaluate the predictive power of such features under experimental conditions that simulate a clinical setting where such applications could potentially find use, which distinguishes our work from previous studies that have been restricted to establishing correlations between different features and bone strength.

We demonstrate our approach in this study by pursuing three different approaches to capturing information pertaining to the trabecular bone microarchitecture in the proximal femur for purposes of complementing conventionally used DXA BMD: (1) statistical moments of the MDCT BMD distribution, (2) morphometric parameters, such as bone fraction, trabecular thickness, etc., and (3) geometrical features derived from the SIM. SIM can be used to extract information related to local geometric properties in point distributions and gray-level patterns.[24,25] Previous work has successfully demonstrated the ability of SIM-derived geometric features to characterize the complex trabecular bone microarchitecture for osteoporosis assessment on different imaging modalities.[19,21,26–29] Once the feature sets are extracted from the trabecular compartment of the femur, they are subsequently processed with different regression models for the prediction task, as discussed in the following sections.

2 Materials and Methods

2.1 Femur Specimens

Femur specimens were harvested from 248 formalin-fixed human cadavers at the Institute of Anatomy at the Ludwig Maximilians University, Munich, Germany, for educational and research purposes, in compliance with local institutional and legislative requirements. Exclusion criteria included (1) identification of diffuse metastatic bone disease or hematologic or metabolic bone disorders other than osteoporosis through histological examination of samples biopsied from the iliac crest and (2) detection of fracture on radiographs or during specimen preparation for storage and scanning. Taking these exclusions into account, a subset of 146 human femur specimens were used in this study. The donors (73 women, 73 men) had a mean life span of 79.39 years (standard deviation: 10.57 years, range: 52 to 100 years). The bones, along with a variable amount of surrounding soft tissue, were removed from the cadavers; the soft tissue was subsequently excised prior to imaging and biomechanical testing. The specimens were degassed for at least 24 h before MDCT. The degassing procedure involved submerging the specimens in a formalin solution within a cylindrical vacuum container, which was subsequently evacuated

to −0.95 bar with a special vacuum pump. During the study, the specimens were stored in fixative solution to prevent storage and air artifacts.

2.2 DXA Measurements

DXA was used to determine BMD in the entire proximal femur, as well as in the neck and trochanter regions. The measurements were performed with a Prodigy Scanner (GE/Lunar; GE Medical Systems, Milwaukee, Wisconsin). The femur specimens were positioned similar to *in vivo* examination conditions: mildly internally rotated in a vessel filled with water to 15 cm in height to simulate soft tissue. The measurements were evaluated by using the Lunar Prodigy Encore 2002 software (GE Medical Systems). It should be noted that DXA measurements were not performed in the femoral head alone, owing to superimposition with the acetabulum in *in vivo* conditions.

2.3 Multidetector CT Measurements

Cross-sectional images of the femora were acquired with a 16-detector CT scanner (Sensation 16; Siemens Medical Solutions, Erlangen, Germany). The specimens were placed in plastic bags filled with a 4% formalin-water solution. These plastic bags were sealed after air was removed by a vacuum pump. These bags were positioned in the scanner to simulate conditions of an *in vivo* examination of the pelvis and proximal femur, with mild internal rotation of the femur. Each specimen was scanned using a protocol involving a collimation and table feed of 0.75 mm and a reconstruction index of 0.5 mm. A high-spatial-resolution reconstruction algorithm (kernel U70u) was used, with a resulting in-plane resolution of $0.29 \times 0.29\ mm^2$. Additional scanning parameters were 120 kVp, 100 mA, an image matrix of 512 × 512 pixels, and a field of view of 100 mm. Voxel size was $0.19 \times 0.19 \times 0.5\ mm^3$. For calibration purposes, a reference phantom with a bone-like and a water-like phase (Osteo Phantom, Siemens Medical Solutions, Erlangen, Germany) was placed in the scanner below the specimens, as shown in Fig. 1.

2.4 Image Processing and Volume of Interest Selection

The outer surface of the femoral cortical shell was segmented in an automated manner using the bone attenuations of the phantom on each image as reference. In 3% of the specimens, the segmentation mask was adversely affected by errors induced by high-grade focal bone loss or penetration of adjacent anatomical structures, such as blood vessels, into the cortex (mostly in the femoral head); manual corrections were performed by one of the two radiologists. An algorithm previously proposed by Huber et al.[22] was used to fit volumes of interest (VOIs) approximated from a sphere for the trabecular region of the femoral head, a cylinder for the neck, and a cone for the greater trochanter. Further details regarding this automated algorithm can be found in Ref. 22. Figure 2 shows the cross-sectional boundary of selected VOIs on different regions of a representative proximal femur specimen. The average number of voxels included in the final mask was 3.77×10^5 for the femoral head, 4.33×10^4 for the neck, and 1.83×10^5 for the trochanter.

2.5 Quantitative CT BMD Measurements

The mean BMD of each VOI was calculated by converting pixel attenuations on MDCT (Hounsfield units) into BMD values (mg/cm^3) using a conversion proposed in Ref. 20. BMD was calculated as

$$\mathrm{BMD} = [HA_B / (HU_B - HU_W)] \cdot (HU - HU_W). \quad (1)$$

Here, HA_W (0 mg/cm^3) and HA_B (200 mg/cm^3) were the densities of the water-like and bone-like parts of the hydroxyapatite calibration phantom, while HU_W and HU_B were their corresponding attenuations on the MDCT images. After BMD conversion, the range of BMD values within the regions of interest (ROIs) were limited to a window of [−200,1200] (values outside this interval were set to the boundary values), which were the calculated BMD values for air and bone according to Eq. (1) in order to emphasize bone content. Examples of ROIs converted to BMD are shown in Fig. 2.

2.6 Biomechanical Tests

The failure load (FL) was assessed with a side-impact test in which a lateral fall on the greater trochanter was simulated, as described previously in Ref. 20. The femur shaft and head were faced downward and could be moved independently of one another while a load was applied to the greater trochanter by using a universal materials testing machine (Zwick 1445; Zwick, Ulm, Germany) with a 10-kN force sensor and dedicated software. FL was defined as the peak of the load-deformation curve. Fracture sites were commonly noted in the neck and trochanter regions for most specimens (trochanter: 54, neck: 66, shaft: 26). The 146 femur specimens used in this study had a mean FL of 3.94 kN (standard deviation: 1.56 kN, range: 0.66 to 8.16 kN). For prediction of relative bone strength, the FL was adjusted for total body height (FL/H) and weight (FL/W), as previously proposed in Ref. 19, and used for subsequent analysis.

2.7 Trabecular Bone Feature Analysis on MDCT

2.7.1 Statistical features—bone mineral density—The MDCT BMD distribution within VOIs was first represented by the following statistical feature sets, referred to in this work as BMD features—(1) mean (BMD.mean), (2) mean and standard deviation (BMD.m2), (3) mean, standard deviation, skewness, and kurtosis (BMD.m4), and (4) 19 quantiles (BMD.p19).

2.7.2 Morphometric parameters—For calculation of morphometric parameters, MDCT images of the femur specimens were first binarized to label the pixels as “bone” or “marrow” for further analysis. For this purpose, a binarizing threshold of 200 mg/cm^3 was previously proposed based on visual examination of images from 30 proximal femur specimens.[20] After binarization, two-dimensional (2-D) parameters were calculated in analogy to standard histomorphometry using the mean intercept length method[30] on every slice of the VOI: bone fraction (bone volume/total volume), trabecular number, trabecular separation, and trabecular thickness. For each parameter, the values calculated on each slice

of a VOI were averaged and then concatenated into a four-dimensional feature vector that represented the morphometric feature set of that VOI.

2.7.3 Geometric features-scaling index method—The SIM[24,25] was used to extract geometrical features that characterized the structural properties of trabecular bone microarchitecture observed in the VOIs. Consider N pixels in an ROI represented by a three-dimensional vector $N_i = (x_i, y_i, g_i)$, $i = \{1, \ldots, N\}$, where x_i and y_i are the spatial dimensions and g_i is the gray-level intensity of the i'th pixel, i.e., $g_i = g(x_i, y_i)$. The application of SIM for a given scale R can be regarded as an image transformation where each pixel of the original ROI is assigned a local scaling property $\alpha_i = \alpha(x_i, R)$ resulting in a transformed SIM ROI $x_i^s = (x_i, y_i, \alpha_i)$. The local scaling property index α can be calculated as

$$\alpha(x_i, R) = \frac{2\sum_{j=1}^{N} (d_{ij}/R)^2 exp\left[-(d_{ij}/R)^2\right]}{\sum_{j=1}^{N} exp\left[-(d_{ij}/R)^2\right]}, \quad (2)$$

where $d_{ij} = \|x_i x_j\|$ is the Euclidean distance between the i'th and the j'th pixel and R defines the Gaussian neighborhood surrounding the i'th pixel. To prevent the ROI border from imposing an artificial local geometry (stemming from pixels close to the border having incomplete neighborhoods within the ROI), α-values were only computed for pixels farther than $2R$ from the border. Such an SIM transformation is computed for every slice in the VOI, and the resulting distribution of α-values reveals nonlinear structural information related to the gray-level patterns in the VOI. The entire distribution of α-values within the VOI was represented by its 19 quantiles, i.e., 5th to 95th. This served as a 19-dimensional geometrical feature vector that characterized the trabecular bone microarchitecture captured within the VOI, as shown in Fig. 3.

We note the absence of an intrinsic scaling factor that defines the relationship between spatial coordinates and intensity that defines each pixel while computing the α-values. We propose to investigate a set of scaling factors (SF) to be applied to the gray-level intensities within the VOI prior to the SIM transformation and choose the optimum SF that minimizes the FL prediction error. In this study, we investigated different values for the scaling factor, i.e., SF = {0.01, 0.1,1,5,10,50,100}, in calculating the scaling index α for each pixel using a set of radii $R = \{1, 2, 3, 4, 5\}$. Examples of applying SIM to VOIs used in this study are shown in Fig. 3.

2.8 Function Approximation

After the calculation of different bone characterizing feature sets, their ability to predict the biomechanical strength of the specimens, as indicated by the FL, was evaluated. Standard multiregression analysis was used for each feature set to assess a linear prediction.

In addition, SVR with a linear kernel[31–33] was used. SVR is an extension of support vector machine (SVM),[34] which is a supervised learning model that is widely used in the literature for different data analysis and pattern recognition tasks. SVM aims to devise a computationally efficient method for identifying a hyperplane that imposes a maximal

margin of separation between two classes of data points in the high-dimension feature space. While many such hyperplanes could exist, the preferred hyperplane (or decision boundary) is the one that has the least chance of misclassifying future data points yet to be encountered. This is determined by maximizing the distance between such a hyperplane and the training points closest to it (known as support vectors) so that the remaining training data are much further from this decision boundary. This key characteristic of SVM, also known as large margin intuition,[35] ensures that the decision boundary is optimally chosen to accommodate unknown data points and likely explains its superiority over other traditional classifiers. Further details of this algorithm can be found in many different resources, including Refs. 32 to 34. While SVMs are used to predict binary labels, SVR allows for the prediction of continuous values and finds use in our study as a function approximation for predicting FL. The SVR implementation was taken from the *lib*SVM library.[36]

2.9 Prediction Performance

The prediction performance of the features was evaluated in a cross-validation strategy where the dataset was divided into a training set (80%) and an independent test set (20%). The independent test set allowed for an unbiased evaluation of the bone strength prediction performance achieved by the regression models. The training set was used to approximate the target function, i.e., FL and adjusted FL (FL/H and FL/W). The resulting model was used to predict the FL (and adjusted FL) of the remaining specimens that constituted the independent test set. The average residual error between the predicted FL (FL_{pred}) and the true FL (FL_{true}) for the test set T_i, $i = 1, \ldots, N_{iter}$, was measured by the root mean square error ($RMSE_{Ti}$).

$$\mathrm{RMSE}_{T_i} = \sqrt{\left\langle (FL_{pred} - FL_{true})^2 \right\rangle_{T_i}}. \quad (3)$$

The calculation of RMSE, as detailed above, was repeated 50 times for different randomly chosen training and test set collections, resulting in an RMSE distribution for each bone feature set. In addition, the coefficient of determination R^2 was calculated to evaluate the correlation between different feature sets and FL_{true} for each iteration. A Wilcoxon signed-rank test was used to compare two RMSE distributions and test for statistical significant differences in performance. The statistical analysis, feature extraction, function approximation, performance evaluation, and significance testing were performed in MATLAB®, version R2010a (MathWorks, Natick, MA).

The overall experimental setup is presented in Fig. 4.

3 Results

3.1 Exploring the Impact of SIM Free Parameters

We were specifically interested in evaluating the impact of the scaling factor SF and radius *R* on the overall performance. Figure 5 shows the impact of different choices of values for SF and *R* on the prediction performance achieved with SIM-derived geometric features and both function approximation models, i.e., multiregression and SVR. The best prediction

performance was observed for SF = 1 when radius $R = 4$ or $R = 5$ was used in the computation of α-values, of which regression model was used. With these settings, the original BMD values were not modified by the scaling factor, while the larger radii allowed better characterization of local structure. Based on these results, SF = 1 was used for further experiments.

3.2 Prediction Performance of Trabecular Bone Descriptors

3.2.1 Femoral head region—The top panel of Fig. 6 shows the performance achieved with different trabecular bone characterizing feature sets extracted from the femoral head in predicting the true FL of the femur specimens. As seen here, the best performance is achieved by BMD.mean, as computed on MDCT, in combination with multiregression (RMSE: 1.11 ± 0.14, R^2: 0.49 ± 0.13). While comparable performance is achieved with MDCT BMD features, such as BMD.m2, all other feature sets are outperformed when multiregression is used. We also note that for the univariate BMD.mean, no difference in performance was achieved regardless of the regression model used.

When SVR is used, the MDCT-extracted BMD.p19 (RMSE: 1.08 ± 0.14, R^2: 0.51 ± 0.13) registers a significant improvement over the MDCT BMD.mean (RMSE: 1.11 ± 0.14, R^2: 0.49 ± 0.13) ($p < 10^{-4}$). Comparable per formance to the MDCT BMD.mean is achieved by the SIM ($R = 5$) feature set (RMSE: 1.13 ± 0.13, R^2: 0.49 ± 0.08) and the feature vector of morphometric parameters (RMSE: 1.12 ± 0.16, R^2: 0.50 ± 0.11).

3.2.2 Femoral neck region—The middle panel of Fig. 6 shows the performance achieved with different trabecular bone characterizing feature sets extracted from the femoral neck in predicting the true FL of the femur specimens. As seen here, the prediction performance of different feature sets are much worse when compared to the results observed in the femoral head.

MDCT-extracted BMD feature sets and morphometric parameters perform better than SIM feature sets when multi-regression is used ($p < 0.05$). While the performance of SIM feature sets significantly improves when SVR is used instead of multiregression ($p < 0.05$), it is still outperformed by MDCT-extracted BMD feature sets and morphometric parameters ($p < 0.05$).

3.2.3 Femoral trochanter region—The bottom panel of Fig. 6 shows the performance achieved with different trabecular bone characterizing feature sets extracted from the femoral trochanter in predicting the true FL of the femur specimens. As seen here, the prediction performance of different feature sets are still worse when compared to the results observed in the femoral head but present an improvement over those observed in the neck.

Again, MDCT-extracted BMD feature sets, such as BMD.m2 and BMD.m4, and morphometric parameters perform better than SIM feature sets regardless of regression model used ($p < 0.001$). We also note that the performance of multivariate feature sets, such as BMD.p19, and SIM feature sets with different R significantly improves when SVR is used in place of multiregression ($p < 0.001$).

3.3 Prediction Performance of Trabecular Bone Descriptors in Combination with DXA BMD

Having identified the femoral head as a suitable candidate region for extracting features that characterize the trabecular bone microarchitecture, we also investigated the prediction performance achieved when BMD measurements from DXA were combined with different trabecular bone descriptors extracted from MDCT images, i.e., BMD.mean, SIM ($R = 5$) features, and morphometric parameters, extracted from the femoral head. These were compared to the prediction performance achieved with DXA BMD alone. The results obtained with multiregression and SVR are shown in Tables 1 and 2, respectively.

As seen in Tables 1 and 2, the best prediction performance for FL is obtained when DXA BMD is combined with SIM-derived geometrical features extracted from the femoral head. Such feature sets achieve significantly better performance than DXA BMD measurements alone ($p < 10^{-4}$). For multivariate feature sets, such as combinations of DXA BMD and SIM, SVR was found to yield significantly better performance at predicting FL over multiregression ($p < 0.01$). Overall, the combination of DXA BMD evaluated from the entire femur and SIM ($R = 5$) features, when processed with SVR, was found to achieve the best performance at predicting FL.

Similar trends in performance were noted when the machine learning task involved predicting FL adjusted for height (FL/H), i.e., the combination of DXA BMD and SIM features extracted from MDCT images of the femoral head were found to yield the best prediction performance. The use of SVR over traditionally used multiregression was also found to contribute to the observed performance. However, the opposite was true when the FL was adjusted for body weight (FL/W), i.e., the combination of DXA BMD and SIM features was significantly worse at predicting FL/W when compared to DXA BMD alone.

4 Discussion

This study presents an automated approach to predicting the biomechanical strength of femur specimens through imaging-based analysis of BMD and local geometry of the trabecular bone microarchitecture. Our results suggest that supplementing conventionally used descriptors of BMD, such as DXA BMD, with geometrical features describing the local structure of the trabecular bone, specifically extracted from the femoral head region, can significantly improve the performance achieved at predicting the FL of such specimens. Furthermore, as the feature sets that characterize the trabecular bone become larger to accommodate such detailed information, SVR is a better choice than multiregression for the machine learning task. The noninvasive approach employed for extracting such predictors from the femoral bone makes them attractive candidates for potential diagnostic biomarkers that could aid in improved prediction of osteoporotic fracture risks, clinical assessment and management of osteoporosis, and quantification of the effects of different therapeutic intervention strategies.

The results of our study are in agreement with other studies involving high-resolution magnetic resonance imaging,[21,25,37,38] MDCT,[18–20,22] and micro CT,[39] which have consistently shown that supplementing DXA BMD with features that characterize trabecular bone microarchitecture can strengthen the corresponding correlation to bone strength.

However, rather than establishing correlations between individual features and bone strength, we propose a new approach where different features are combined to form high-dimensional feature vectors, which are subsequently used to train and test supervised learning algorithms for predicting bone strength. As seen in our study, this combination of feature extraction and machine learning exhibits improved prediction performance and easily extends to computer-aided diagnosis applications in a clinical setting. The modular design of our approach, as detailed in Fig. 4, allows easy integration of different feature extraction or machine learning techniques, which could precipitate further improvements in prediction performance.

Our results suggest that when features are extracted from different regions of the proximal femur on MDCT, the best prediction performance is achieved when features are extracted from the head region of the proximal femur, as seen in Fig. 6. This observation is in accordance with other studies that have also shown that bone strength is highly correlated to BMD measurements performed in the femoral head region.[19,20,22] We suspect that such findings suggest that the femoral head is an important load-bearing region in the hip joint, and thus, its features are directly correlated with the biomechanical strength of the proximal femur as a whole. We also note that features extracted from the trochanter region exhibit the second-best prediction performance, while those extracted from the neck region perform the worst. This trend is also observed in other studies that have investigated correlation between bone strength and BMD measurements extracted from the head, neck, and trochanter regions.[19,20,22] The observed results seem to correspond with the volume of the VOIs extracted, which was largest in the femoral head and smallest in the neck. However, we found no significant differences in correlation between the FL of the specimens and the volumes of VOIs extracted from the three anatomical regions. In addition, the features investigated in this study, such as SIM-derived geometrical features, are not directly reliant on VOI size, and the VOIs from all regions are sufficiently large to rule out poor data sampling of voxels during the feature extraction process. Thus, any contribution of the VOI volume to the prediction performance is expected to be minimal at best.

Interestingly, while BMD features and morphometric parameters extracted from the femoral head on MDCT outperformed SIM features when evaluated separately, they did not contribute to improvements in prediction performance when combined with DXA BMD measurements. In fact, the best performance was noted when DXA BMD was combined with SIM features from the femoral head, as seen in Tables 1 and 2. This suggests that these SIM features, which capture local geometrical properties of the trabecular bone microarchitecture, provide information that is complementary to that captured by DXA BMD measurements in the proximal femur. However, such SIM features likely do not capture adequate global information about the trabecular bone structure to be as effective as the MDCT BMD features or morphometric parameters when used independently. Similar trends are noted when the machine learning task involved predicting FL adjusted for height (FL/H). However, as seen in Tables 1 and 2, the performance of such features is worse when predicting FL adjusted for body weight (FL/W). This is not surprising given that previous studies have noted weak correlations between features characterizing trabecular bone structure and FL/W owing to the strong influence that body weight has on FL.[19]

Finally, we also investigated the impact of replacing multiregression with SVR for the machine learning task. Our results suggest that the prediction performance achieved by the feature sets is not affected by the choice of regression model when the feature sets in question involve a small number of features (e.g., most BMD-derived feature sets). However, as the feature set becomes larger, SVR is clearly better equipped to appropriately process the input feature sets and hence outperforms multiregression with such feature sets (e.g., SIM feature sets, combined DXA BMD and SIM feature sets), as seen in Fig. 6 and Tables 1 and 2.

We acknowledge certain limitations with the experimental setup of the current study. The femur specimens used in our analysis were not scanned *in situ* but in a water bath as a relatively simple model to simulate the soft tissue environment. *In vivo*, soft tissue scatter may compromise the image quality compared to the *in vitro* setting of this study.[40] Also, specimens were harvested from formalin-fixed cadavers, which may have affected their biomechanical properties. Given that the MDCT images were not acquired with iso-tropic voxels, both SIM features and histomorphometric parameters were extracted in 2-D from every slice of the VOI. We also note that the geometric features used in this study do not take into account the intrinsic anisotropy of the trabecular bone compartment. Recent studies have shown that anisotropic geometric features can provide limited improvements in achieved correlation with bone strength.[41] Our study was focused on investigating the ability of features that characterize the trabecular bone for predicting FL and only used previously proposed anisotropy-independent geometric features. However, the approach proposed here can be extended to include anisotropy-weighted features as well, and will be explored in future studies.

5 Conclusion

This study proposes an automated approach to predicting biomechanical bone strength in proximal femur specimens through characterization of BMD and trabecular bone microarchitecture used in conjunction with machine learning techniques. The results of our work suggest that supplementing conventionally used DXA BMD measurements with SIM-derived geometrical features that characterize the trabecular compartment in the femoral head for generating prediction models with SVR can effect significant improvements in bone strength prediction performance. The automated and objective manner in which the trabecular bone microarchitecture is analyzed and the subsequent rediction performance achieved suggest tremendous potential in serving as diagnostic biomarkers for osteoporosis diagnosis, tracking disease progression, and evaluating responses to therapeutic intervention.

However, trabecular bone structure analysis using high-resolution MDCT at the proximal femur is currently not feasible for osteoporosis diagnostics and therapy monitoring in clinical practice owing to its radiation dose equivalent lying in the upper range of medically acceptable radiation exposure. Future studies will need to overcome this drawback before larger controlled trials can be conducted for validating the applicability of our approach in a clinical setting.

Acknowledgments

This research was funded in part by the National Institutes of Health Award R01-DA-034977, the Clinical and Translational Science Award 5-28527 within the Upstate New York Translational Research Network of the Clinical and Translational Science Institute, University of Rochester, and by the Center for Emerging and Innovative Sciences, a NYSTAR-designated Center for Advanced Technology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank Dr. Matthias Priemel (Department of Trauma, Hand and Reconstructive Surgery, University Medical Center Hamburg-Eppendorf, Germany) for performing histology examinations of the iliac crest biopsies.

References

1. Pietschmann P, et al. Osteoporosis: an age-related and gender-specific disease—a mini-review. Gerontology. 2009; 55(1):3–12. [PubMed: 18948685]
2. Cooper C, Campion G, Melton LJ. Hip fractures in the elderly: a world-wide projection. Osteoporosis Int. 1992; 2(6):285–289.
3. Black DM, et al. An assessment tool for predicting fracture risk in postmenopausal women. Osteoporosis Int. 2001; 12(7):519–528.
4. Kanis JA, et al. Assessment of fracture risk. Osteoporosis Int. 2005; 16(6):581–589.
5. Boehm H, et al. Improved performance of hip DXA using a novel region of interest in the upper part of the femoral neck: in vitro study using bone strength as a standard of reference. J. Clin. Densitom. 2005; 8(4):488–494. [PubMed: 16311437]
6. Mazess R, et al. Performance evaluation of a dual-energy x-ray bone densitometer. Calcif. Tissue Int. 1989; 44(3):228–232. [PubMed: 2493329]
7. Eckstein F, et al. Bone strength at clinically relevant sites displays substantial heterogeneity and is best predicted from site-specific bone densitometry. J. Bone Miner. Res. 2002; 17(1):162–171. [PubMed: 11771664]
8. Johnell O, et al. Predictive value of BMD for hip and other fractures. J. Bone Miner. Res. 2005; 20(7):1185–1194. [PubMed: 15940371]
9. Bolotin HH, Sievanen H, Grashuis JL. Patient- specific DXA bone mineral density inaccuracies: quantitative effects of non-uniform extra osseous fat distributions. J. Bone Miner. Res. 2003; 18(6): 1020–1027. [PubMed: 12817754]
10. Li W, et al. Automated registration of hip and spine for longitudinal quantitative CT studies: integration with 3D densitometric and structural analysis. Bone. 2006; 38(2):273–279. [PubMed: 16199215]
11. Lang TF, et al. Measurement of bone mineral density at the spine and proximal femur by volumetric quantitative computed tomography and dual-energy x-ray absorptiometry in elderly women with and without vertebral fractures. Bone. 2002; 30(1):247–250. [PubMed: 11792593]
12. Lang TF, et al. Volumetric quantitative computed tomography of the proximal femur: precision and relation to bone strength. Bone. 1997; 21(1):101–108. [PubMed: 9213015]
13. Bousson V, et al. Volumetric quantitative computed tomography of the proximal femur: relationships linking geometric and densitometric variables to bone strength role for compact bone. Osteoporosis Int. 2006; 17(6):855–864.
14. Genant HK, et al. Quantitative computed tomography in assessment of osteoporosis. Semin. Nucl. Med. 1987; 17(4):316–333. [PubMed: 3317846]
15. Taylor BC, et al. Long-term prediction of incident hip fracture risk in elderly white women: study of osteoporotic fractures. J. Am. Geriatr. Soc. 2004; 52(9):1479–1486. [PubMed: 15341549]
16. Lang T, et al. Cortical and trabecular bone mineral loss from the spine and hip in long-duration spaceflight. J. Bone Miner. Res. 2004; 19(6):1006–1012. [PubMed: 15125798]
17. Adams JE. Quantitative computed tomography. Eur. J. Radiol. 2009; 71(3):415–424. [PubMed: 19682815]
18. Huber MB, et al. Predicting the biomechanical strength of proximal femur specimens with bone mineral density features and support vector regression. Proc. SPIE. 2012; 8315:83151W.

19. Baum T, et al. Automated 3D trabecular bone structure analysis of the proximal femur–prediction of biomechanical strength by CT and DXA. Osteoporosis Int. 2010; 21(9):1553–1564.
20. Bauer JS, et al. Structural analysis of trabecular bone of the proximal femur using multislice computed tomography: a comparison with dual x-ray absorptiometry for predicting biomechanical strength in vitro. Calcif. Tissue Int. 2006; 78(2):78–89. [PubMed: 16467973]
21. Huber MB, et al. Prediction of biomechanical properties of trabecular bone in MR images with geometric features and support vector regression. IEEE Trans. Biomed. Eng. 2011; 58(6):1820–1826. [PubMed: 21356612]
22. Huber MB, et al. Proximal femur specimens: automated 3D trabecular bone mineral density analysis at multi detector CT-correlation with biomechanical strength measurement. Radiology. 2008; 247(2):472–481. [PubMed: 18430879]
23. Huber MB, et al. Proximal femur specimens: automated 3D trabecular bone mineral density analysis at multi detector CT-correlation with biomechanical strength measurement. Radiology. 2008; 247(2):472–481. [PubMed: 18430879]
24. Raeth C, et al. Analysing large-scale structure—I. Weighted scaling indices and constrained randomization. MNRAS. 2002; 337(2):413–421.
25. Jamitzky F, et al. Scaling-index method as an image processing tool in scanning-probe microscopy. Ultramicroscopy. 2001; 86(1–2):241–246. [PubMed: 11215629]
26. Boehm HF, et al. Local 3D scaling properties for the analysis of trabecular bone extracted from high-resolution magnetic resonance imaging of human trabecular bone. Invest. Radiol. 2003; 38(5):269–280. [PubMed: 12750616]
27. Raeth C, et al. Improving the textural characterization of trabecular bone structure to quantify its changes: the locally adapted scaling vector method. Proc. SPIE. 2005; 5747:240–248.
28. Mueller D, et al. The 3D-based scaling index algorithm: a new structure measure to analyze trabecular bone architecture in high-resolution MR images in vivo. Osteoporosis Int. 2006; 17(10):1483–1493.
29. Huber MB, et al. Development and testing of texture discriminators for the analysis of trabecular bone in proximal femur radiographs. Med. Phys. 2009; 36(11):5089–5098. [PubMed: 19994519]
30. Parfitt M, et al. Bone histomorphometry: standardization of nomenclature, symbols and units. Report of the ASBMR histomorphometry nomenclature committee. J. Bone Miner. Res. 1987; 2(6):595–610. [PubMed: 3455637]
31. Drucker, H., et al. Advances in Neural Information Processing Systems. Vol. Vol. 9. MIT Press; Cambridge, Massachusetts: 1997. Support vector regression machines; p. 155-161.
32. Schoelkopf B, et al. New support vector algorithms. Neural Comput. 2000; 12(5):1207–1245. [PubMed: 10905814]
33. Duda, RO.; Hart, PE.; Stork, DG. Pattern Classification. Wiley-Interscience Publication; New York: 2000.
34. Cortes C, Vapnik V. Support-vector networks. Mach. Learn. 1995; 20(3):273–297.
35. Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press; Cambridge, Massachusetts: 2000.
36. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2011; 2(3):27.
37. Monetti RA, et al. Scaling index method: a novel nonlinear technique for the analysis of high-resolution MRI of human bones. Proc. SPIE. 2003; 5032:1777–1786.
38. Carballido-Gamio J, et al. Characterization of trabecular bone structure from high-resolution magnetic resonance images using fuzzy logic. Magn. Reson. Imaging. 2006; 24(8):1023–1029. [PubMed: 16997072]
39. Ollivier M, et al. Radiographic bone texture analysis is correlated with 3D microarchitecture in the femoral head, and improves the estimation of the femoral neck fracture risk when combined with bone mineral density. Eur. J. Radiol. 2013; 82(9):1494–1498. [PubMed: 23756323]
40. Bauer J, et al. Analysis of trabecular bone structure with multi detector spiral computed tomography in a simulated soft-tissue environment. Calcif. Tissue Int. 2007; 80(6):366–373. [PubMed: 17520165]

41. Raeth C, et al. Strength through structure: visualization and local assessment of the trabecular bone structure. New J. Phys. 2008; 10(12):125010.

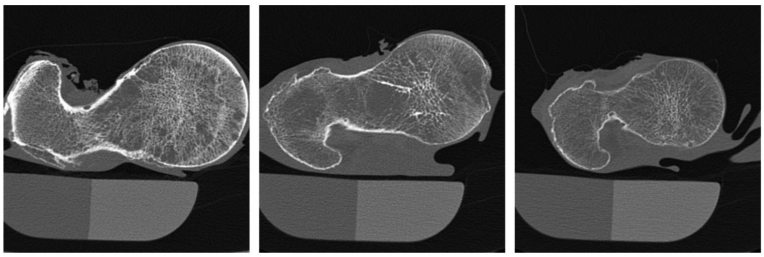

Fig. 1.

Multidetector computer tomography (MDCT) images of selected femur specimens. From left to right, the specimens are categorized as high, medium, and low, based on failure load. The osteo phantom used for each specimen is also shown at the bottom.

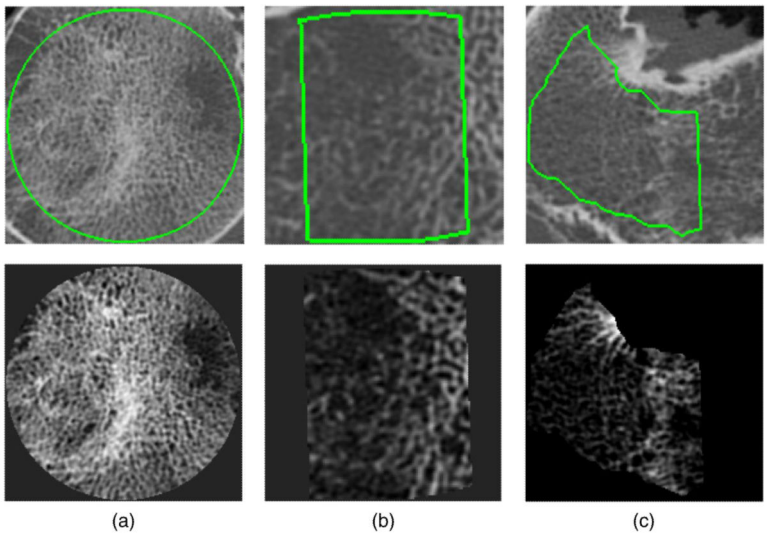

Fig. 2.

Cross-sectional regions of interest (ROIs) of the volumes of interest (VOIs) defined in the femoral head (a), neck (b), and trochanter (c). The top row shows the ROIs overlaid on MDCT images of these regions. The bottom row shows these ROIs where pixel intensities are indicative of bone mineral density (BMD, mg/cm^3). Note that the neck and trochanter images are of different scales and, thus, zoomed in for purposes of presentation only.

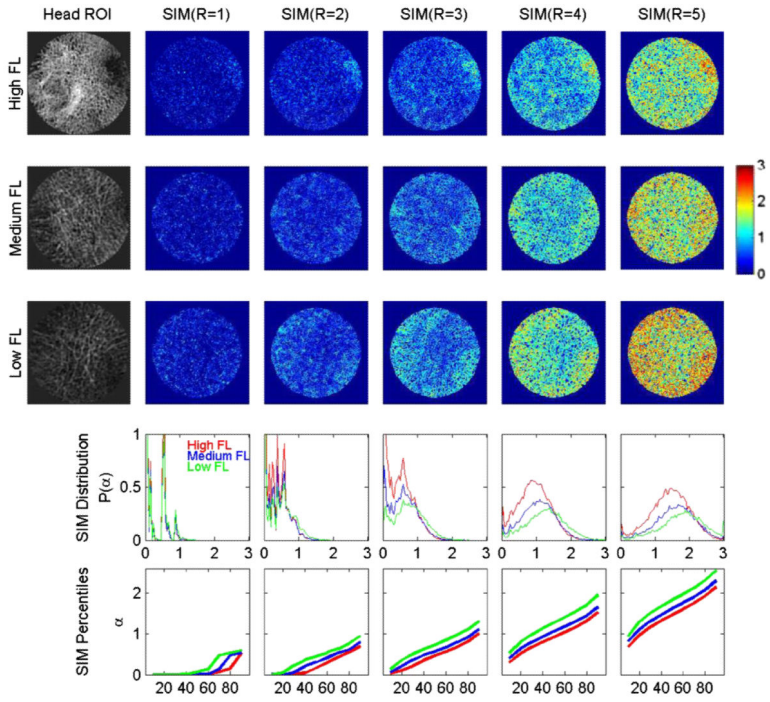

Fig. 3.

ROIs with BMD values extracted from the head VOIs of three specimens characterized as having high (8.16 kN, first row), medium (4.07 kN, second row), and low (1.15 kN, third row) failure load, respectively, are shown on the leftmost column of the first three rows. Subsequent columns in these rows show the scaling index method (SIM) transformations achieved with different radii, $R = \{1, 2, 3, 4, 5\}$ and scaling factor, SF = 1. Scaling indices (α) are color-coded according to the colorbar shown on the right. The (α) histograms for these ROIs for different SIM radii are shown in the fourth row; representations of these histograms using 19 quantiles are shown in the fifth row. Note that only the central slice of each VOI is shown here; however, the histograms and quantile curves represent the distribution of α-values within the entire VOI. As seen here, the histograms and quantile curves can be used to distinguish between femur specimens of different biomechanical strengths.

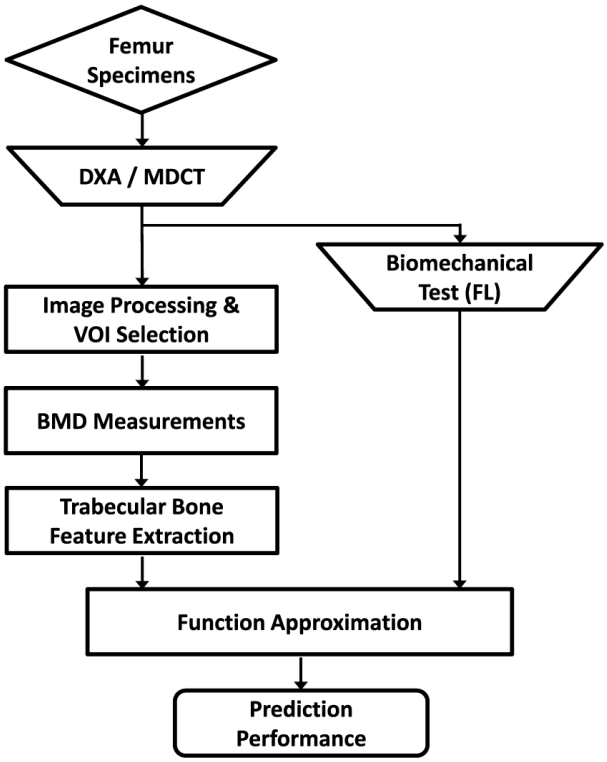

Fig. 4.

Overview of the experimental setup and methods used. The true failure load (FL_{true}) was recorded from biomechanical tests after MDCT imaging. Images were then postprocessed to facilitate conversion of intensity values from Hounsfield units to BMD. Different trabecular bone features (statistical moments of MDCT BMD distribution, geometrical features from SIM, and morphometric parameters) were computed from VOIs annotated on the postprocessed MDCT images. Two function approximation methods, i.e., multiregression and support vector regression, were then used to predict FL_{true}. Prediction performance was quantified using root mean square error (RMSE) and R^2.

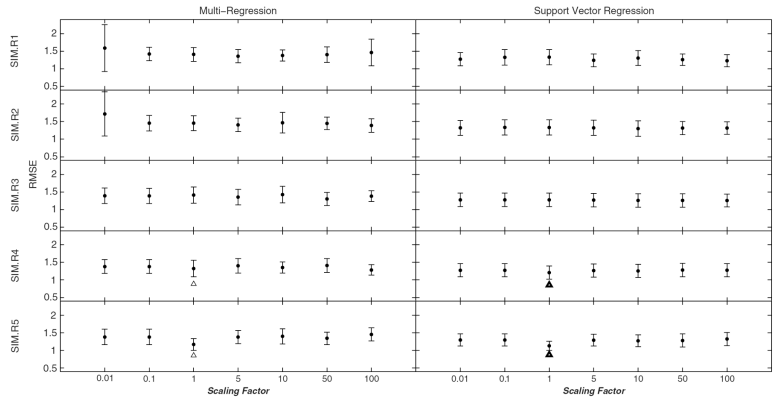

Fig. 5.

Prediction performance (RMSE) achieved with SIM-derived geometric features computed for different radii (*R*) and scaling factor (SF) values using both multiregression and support vector regression. Each RMSE distribution is represented by the central mark that corresponds to the median, and edges that correspond to the 25th and 75th percentile. As seen here, the best prediction performance is achieved for SF = 1 and *R* = 5, as marked with bold triangles.

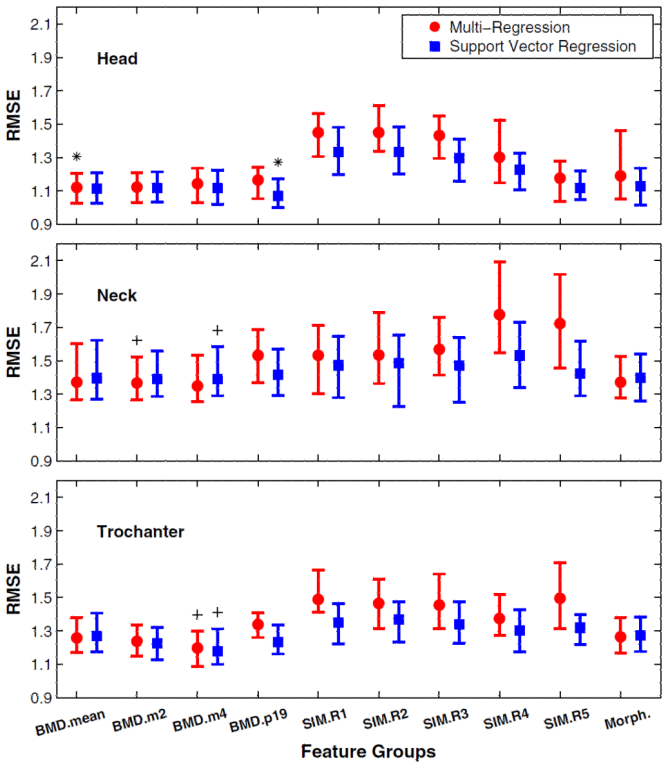

Fig. 6.

Comparison of prediction performance (RMSE) for trabecular bone characterizing features extracted from the femoral head (top), neck (middle), and trochanter (bottom) on MDCT, i.e., mean BMD, SIM features, and morphometric parameters (Morph.), when processed with both multiregression and support vector regression. For each RMSE distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. The best prediction performance was noted achieved with both regression models in the head (marked with *) when compared to the neck and trochanter (best features marked with +).

Table 1

Prediction performance [mean root mean square error (RMSE) (std)] achieved from multiregression by combining dual-energy x-ray absorptiometry (DXA) bone mineral density (BMD) measured from different regions with different feature sets extracted from the femoral head region. Note that "total" indicates that the DXA BMD measurement was determined from the entire proximal femur. "–" indicates that no multidetector computer tomography (MDCT)-derived features were used. As seen here, combination of scaling index method (SIM) features from head region with DXA BMD yields significantly better prediction performance as opposed to using DXA BMD alone.

DXA BMD	MDCT features from femoral head	Failure load (FL) (kN)	FL adjusted for total body height (FL/H) (kN/cm)	FL adjusted for total body weight (FL/W) (kN/kg)
Total	–	0.95 (0.12)	5.37×10^{-3} (7.04×10^{-4})	1.84×10^{-2} (1.68×10^{-3})
	BMD.mean	0.96 (0.12)	5.44×10^{-3} (6.92×10^{-4})	1.86×10^{-2} (1.62×10^{-3})
	SIM.R5	**0.90 (0.10)**	**5.29×10^{-3} (5.78×10^{-4})**	1.97×10^{-2} (2.80×10^{-3})
	Morphometric parameters (Morph.)	0.95 (0.15)	5.44×10^{-3} (8.55×10^{-4})	1.82×10^{-2} (1.78×10^{-3})
Neck	–	0.97 (0.12)	5.52×10^{-3} (6.86×10^{-4})	1.85×10^{-2} (1.77×10^{-3})
	BMD.mean	0.98 (0.12)	5.54×10^{-3} (6.59×10^{-4})	1.87×10^{-2} (1.75×10^{-3})
	SIM.R5	**0.94 (0.11)**	5.55×10^{-3} (6.20×10^{-4})	1.98×10^{-2} (2.75×10^{-3})
	Morph.	0.97 (0.14)	5.55×10^{-3} (7.96×10^{-4})	1.84×10^{-2} (1.94×10^{-3})
Trochanter	–	0.94 (0.12)	5.45×10^{-3} (7.22×10^{-4})	1.87×10^{-2} (1.68×10^{-3})
	BMD.mean	0.96 (0.12)	5.54×10^{-3} (7.07×10^{-4})	1.89×10^{-2} (1.62×10^{-3})
	SIM.R5	**0.90 (0.10)**	**5.34×10^{-3} (6.26×10^{-4})**	1.99×10^{-2} (2.87×10^{-3})
	Morph.	0.94 (0.15)	5.47×10^{-3} (8.63×10^{-4})	1.82×10^{-2} (1.86×10^{-3})

Note: The significance of bold values is $p < 10^{-3}$.

Table 2

Prediction performance [mean RMSE (std)] achieved from support vector regression by combining DXA BMD measured from different regions with different feature sets extracted from the femoral head region. Note that "total" indicates that the DXA BMD measurement was determined from the entire proximal femur. "–" indicates that no MDCT-derived features were used. As seen here, combination of SIM features from head region with DXA BMD yields significantly better prediction performance than using DXA BMD alone.

DXA BMD	MDCT features from femoral head	FL (kN)	FL/H (kN/cm)	FL/W (kN/kg)
Total	–	0.95 (0.12)	5.42×10^{-3} (7.12×10^{-4})	1.84×10^{-2} (1.69×10^{-3})
	BMD.mean	0.96 (0.11)	5.47×10^{-3} (6.90×10^{-4})	1.87×10^{-2} (1.64×10^{-3})
	SIM.R5	**0.87 (0.12)**	**5.14×10^{-3} (6.96×10^{-4})**	1.93×10^{-2} (2.10×10^{-3})
	Morph.	0.95 (0.14)	5.56×10^{-3} (8.30×10^{-4})	1.86×10^{-2} (1.90×10^{-3})
Neck	–	0.97 (0.12)	5.52×10^{-3} (6.81×10^{-4})	1.85×10^{-2} (1.72×10^{-3})
	BMD.mean	0.97 (0.12)	5.53×10^{-3} (6.58×10^{-4})	1.88×10^{-2} (1.75×10^{-3})
	SIM.R5	**0.88 (0.13)**	**5.30×10^{-3} (7.40×10^{-4})**	1.96×10^{-2} (2.42×10^{-3})
	Morph.	0.96 (0.13)	5.52×10^{-3} (7.75×10^{-4})	1.87×10^{-2} (2.03×10^{-3})
Trochanter	–	0.94 (0.12)	5.49×10^{-3} (7.31×10^{-4})	1.86×10^{-2} (1.70×10^{-3})
	BMD.mean	0.98 (0.13)	5.61×10^{-3} (7.02×10^{-4})	1.89×10^{-2} (1.65×10^{-3})
	SIM.R5	**0.89 (0.12)**	**5.24×10^{-3} (6.73×10^{-4})**	1.98×10^{-2} (2.06×10^{-3})
	Morph.	0.97 (0.15)	5.63×10^{-3} (8.74×10^{-4})	1.87×10^{-2} (1.97×10^{-3})

Note: The significance of bold values is $p < 10^{-4}$.