BMES BIOMEDICAL ENGINEERING SOCIETY

Original Article

3D Analysis of the Proximal Femur Compared to 2D Analysis for Hip Fracture Risk Prediction in a Clinical Population

FATEMEH JAZINIZADEH[1] and CHERYL E. QUENNEVILLE[1,2]

[1]Department of Mechanical Engineering, McMaster University, ABB-C308, 1280 Main St. West, Hamilton, ON L8S 4L8, Canada; and [2]School of Biomedical Engineering, McMaster University, Hamilton, ON, Canada

Associate Editor Lyndia (Chun) Wu oversaw the review of this article.

Abstract—Due to the adverse impacts of hip fractures on patients' lives, it is crucial to enhance the identification of people at high risk through accessible clinical techniques. Reconstructing the 3D geometry and BMD distribution of the proximal femur could be beneficial in enhancing hip fracture risk predictions; however, it is associated with a high computational burden. It is also not clear whether it provides a better performance than 2D model analysis. Therefore, the purpose of this study was to compare the 2D and 3D model reconstruction's ability to predict hip fracture risk in a clinical population of patients. The DXA scans and CT scans of 16 cadaveric femurs were used to create training sets for the 2D and 3D model reconstruction based on statistical shape and appearance modeling. Subsequently, these methods were used to predict the risk of sustaining a hip fracture in a clinical population of 150 subjects (50 fractured, and 100 non-fractured) that were monitored for five years in the Canadian Multicentre Osteoporosis Study. 3D model reconstruction was able to improve the identification of patients who sustained a hip fracture more accurately than the standard clinical practice (by 40%). Also, the predictions from the 2D statistical model didn't differ significantly from the 3D ones ($p > 0.76$). These results indicated that to enhance hip fracture risk prediction in clinical practice implementing 2D statistical modeling has comparable performance with lower associated computational load.

Keywords—Hip fracture risk, Proximal femur, DXA scanning, 2D statistical modeling, 3D statistical modeling.

Address correspondence to Cheryl E. Quenneville, Department of Mechanical Engineering, McMaster University, ABB-C308, 1280 Main St. West, Hamilton, ON L8S 4L8, Canada. Electronic mail: quennev@mcmaster.ca

INTRODUCTION

Osteoporosis is a disease most common in older adults, which results in low bone mass and micro-architectural deterioration, and can lead to a bone fracture.[31] The hip (proximal femur) is one of the most common sites affected by osteoporosis, the fracture of which can result in severe morbidity and mortality.[12,27] Patients with an early diagnosis of osteoporosis can benefit from protective measures (e.g., targeted exercise, hip protectors, pharmaceutical interventions) to prevent these fractures.[25,39] Currently, the most common method for the diagnosis of osteoporosis relies on the measurement of bone mineral density (BMD) from a dual-energy X-ray absorptiometry (DXA) scan.[31] However, studies have shown that the DXA scan alone is not sufficient in identifying all patients at high risk of sustaining a hip fracture,[20,29] and fifty percent of hip fractures occur in patients with non-osteoporotic DXA scans.[9,36]

DXA scans mainly measure the average BMD in certain regions of the bone, from which the mechanical properties of the bone can be inferred; however, the strength of a femur depends on its geometry,[8,16] BMD distribution pattern,[40,42] and trabecula's quality [17,32] as well. Also, the probability of sustaining a fall and the variability of the loads applied to an individual during a fall are other factors that have been neglected by this approach.[22] Some studies have tried to incorporate these factors in fracture risk assessments to enhance the identification of patients at a higher risk of sustaining a fracture.[1,3,10,15] Considering the effect of a femur's geometry and BMD distribution can be done in 2D using DXA scans and X-ray radiographs, or in 3D using Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI).[13] While 3D imaging pro-

vides more insight into the whole geometry and density distribution of the bone, it is not always feasible to use 3D imaging, due to the expense, time, accessibility, and radiation levels. Therefore, it is anticipated that 2D imaging (DXA scans) will remain as the primary method of diagnosing osteoporosis and consequently fracture risk.[41]

To enhance hip fracture risk prediction, researchers have performed 2D analysis on medical images (either DXA scan or another X-ray based radiography of hip) and their results have shown that it has noticeable improvements over BMD alone (up to 42% improvement in identifying patients at high risk).[1,6,13] Also, to gain the benefits of 3D imaging, other studies have tried to develop 3D structures from 2D scans, using statistical modeling.[10,35] This method allows inference of both geometry and architecture of bones in 3D based on a template model that is created from a training set. Reconstruction of the 3D model of the proximal femur based on a 2D DXA image can enable estimation of the 3D features that otherwise cannot be evaluated in a 2D image.[11] The generated 3D model can also be used for further numerical analysis such as finite element analysis.[7] Some studies have investigated hip fracture risk by considering the effect of the femur's shape and BMD distribution through 3D statistical models, and their results showed that fracture risk estimation was substantially improved compared to using traditional BMD evaluation (up to 45% improvement in identifying patients at high risk).[3,37]

While generating 3D models might be a necessity in further numerical analysis, it is not completely clear if recreating the 3D model from a 2D image to only investigate the geometry and BMD distribution pattern in the femur will have an advantage over investigating the geometry and BMD distribution pattern in 2D alone. Since the 2D and 3D model studies to estimate hip fracture risk were performed based on different training sets and testing groups, the potential to do any direct comparison between them is limited. Therefore the aims of this study were (1) to create 3D shape and BMD distribution models of the proximal femur based on DXA scans, (2) investigate the accuracy of the proposed 3D model reconstruction in comparison to CT scans, and (3) apply 2D and 3D model analysis methods to a clinical population to estimate their hip fracture risk and compare it to their fracture history in a five-year period after the baseline.

MATERIAL AND METHODS

This study had two phases: in phase one, the 2D and 3D analyses were developed using cadaveric specimens (which had 2D and 3D images). In phase two, the techniques were tested and evaluated on a clinical population who had 2D images and fracture history over five years. Ethics approval was granted through McGill University and the appropriate ethics review boards for each participating center. All participants gave written informed consent.

Sixteen isolated cadaveric femurs with no report of any musculoskeletal disease (8 male 61.6 ± 10 years old, 8 female 64.7 ± 6.6 years old) were used for the training sets in this study.[14] Each femur was scanned with a DXA scanner (Hologic Discovery A, Hologic, Inc., Marlborough, MA, USA) and a CT scan machine (GE LightSpeed, GE Healthcare, Chicago, Illinois, USA) with 0.625 mm slice thickness, 0.7 mm in plane resolution, and 120 kV tube voltage, to obtain the geometry and 2D areal and 3D spatial BMD distribution within the bone.

3D Model Reconstruction from DXA Scan

Image processing was performed using MATLAB Image Processing Toolbox (MATLAB R2019b, MathWorks, Natick, Massachusetts, US). Reconstruction of the model consisted of two stages: (1) creating the BMD and geometry template models, where the 3D template models were created and the main modes of variation in the geometry and BMD distribution in the training set were found, and (2) assessing a new scan, where each new DXA scan can be described by the template model plus some variation from it based on the calculated main modes of variation from the first step. The values of these modes were estimated through an optimization process to minimize the differences between the calculated model and the real DXA scan.

Creating the BMD and Geometry Template Models

To create the 3D statistical shape and appearance model, the CT scans of the cadaveric femurs were used to generate 3D models for the training set (MIMICS 22.0, Materialise NV, Leuven, Belgium). For each scan, an STL file was generated to represent the geometry of the proximal femur and a voxel-based mesh was created to describe the BMD distribution in the bones. Twenty-seven geometric landmarks were assigned to each of the models (Fig. 1). The landmarks were placed on the exterior surface of the bone and were based on the identifiable anatomical features. After aligning and removing the effect of translation, rotation, and scaling (using General Procrustes Analysis, GPA) the average landmark coordinates were calculated. Then all models were warped to the average landmark coordinates. The minimum number of vertices from the CT scan 3D model creation was 2255
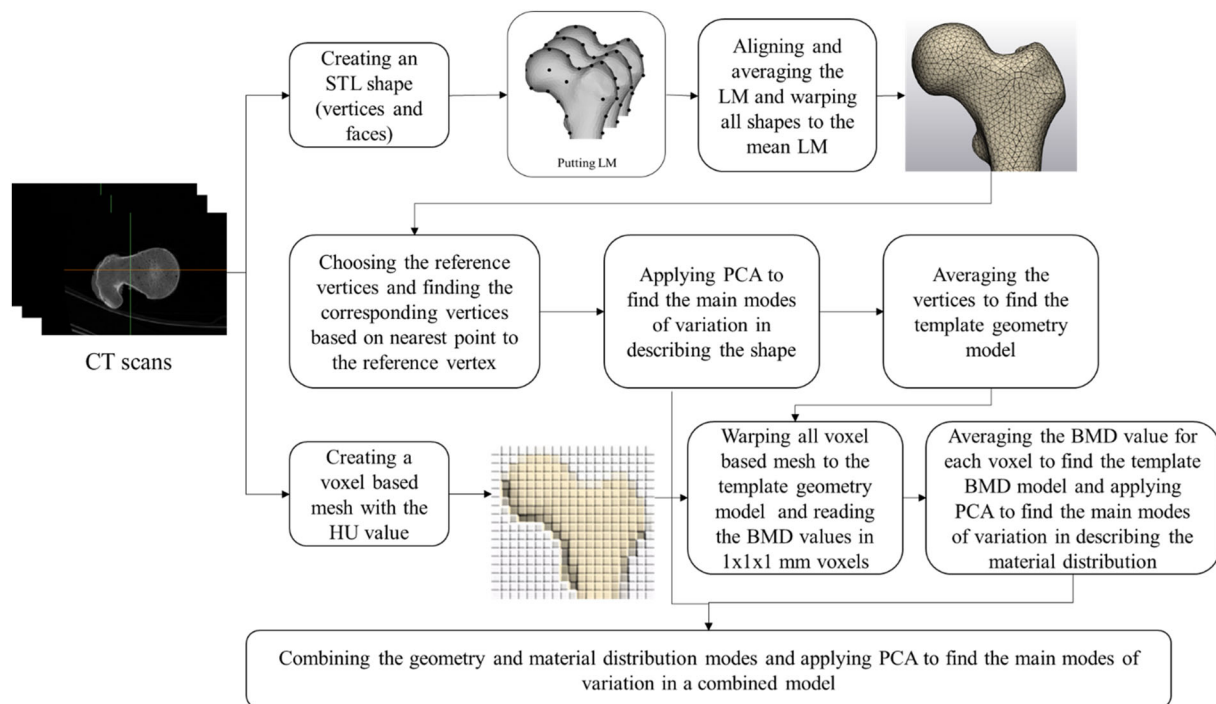
FIGURE 1. Flowchart of creating the 3D statistical shape and appearance models. From the CT scans of isolated cadaveric femurs an STL file to show the surface geometry and a voxel-based mesh to show the BMD distribution was generated. *LM* landmarks, *PCA* principal component analysis, *HU* hounsfield unit, *BMD* bone mineral density.

vertices, so these were chosen as the reference vertices and corresponding vertices in other 3D models were selected automatically by a closest point algorithm. The average 3D shape was thus calculated (creating the template geometry model), and then all 3D models as well as the voxel-based mesh were warped to the average model. Hounsfield Unit (HU) values were then captured in $1 \times 1 \times 1$ mm voxels for each warped 3D model and normalized to the mean and standard deviation of that model. They were then averaged for all specimens to create the template BMD model. Finally, Principal Component Analysis (PCA) was used on both geometry and BMD data to find the main modes of variation in them, which were then gathered in a matrix and PCA was used again to find the main modes of variation for statistical shape and appearance model combined. In describing the geometry, these main modes could correspond to various geometry traits (e.g., the neck-shaft angle, neck length, …), and in describing the BMD distribution, they correspond how much density is concentrated in various regions (cortical thickness, density in the trochanteric area compared to the femoral head,…). More examples with illustration can be found in a previous study.[13]

Assessing a New Scan

To create the 3D model of each femur from its DXA scan, 19 landmarks were assigned on the contour of the femur. Next, the geometry template model was adjusted by changing the weight of its main modes of variation to minimize the difference between the DXA scan and anterior-posterior projection of the 3D model (Fig. 2). After estimating the geometry modes, the femur's shape from the 2D DXA scan was warped to the anterior-posterior projection of the 3D geometry template, and then the gray value of each pixel was captured and normalized to the mean and standard deviation of all pixels for that scan. To obtain the anterior-posterior projection of the 3D template model, the intensity of the voxels (representing the BMD) along the sagittal axis were accumulated to find each pixel's intensity in the 2D image (anterior-posterior projection). The intensity of each pixel was then normalized to the mean and standard deviation of all pixels in that image (the 2D projection).

The 3D BMD template model was changed by altering the weight of its modes, and in each iteration, the anterior-posterior projection of the adjusted template was compared to the warped DXA scan to minimize the differences (pixels' intensity) between the two and eventually finding the weight of the BMD modes through an optimization process. In the end, both the calculated geometry and BMD modes were combined to find the statistical shape and appearance modes.
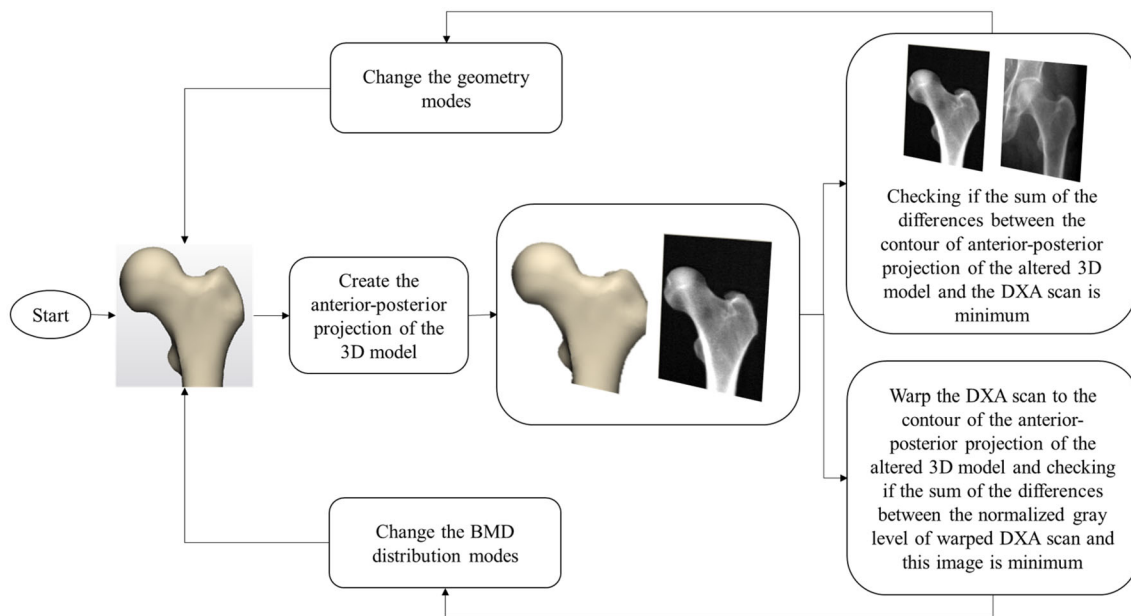
FIGURE 2. The flowchart of finding the modes for a new DXA scan. The modes are found through an optimization process to minimize the difference between the anterior-posterior projection of the altered template model and the DXA scan.

Evaluation of the 3D Model Reconstruction

To evaluate the accuracy of the 3D model reconstruction, the leave-one-out cross-validation technique[13,35] was used on the 16 cadaveric specimens. So, to create the 3D model of each femur from its DXA scan, the CT scans of the other 15 specimens were used in the training set to create the template models and find the main modes of variations. After reconstructing the 3D model for each femur, the created 3D models were compared to the CT-based 3D models. This was evaluated based on the minimum point to surface distance between each vertex from the 3D model reconstruction and the 3D model from the CT scan, as well as the BMD values of the corresponding voxels.

Clinical Data

The subjects used in this study were recruited by the Canadian Multicentre Osteoporosis Study (CaMos). A total of 150 patients' data was used (Table 1), 50 of whom sustained a hip fracture within five years of the baseline DXA scan with a Hologic DXA scanner (Hologic, Inc, Marlborough, MA).

Predicting the Fracture Risk Based on 3D Model Reconstruction

In the clinical application, to create the 3D model of each subject's proximal femur from its DXA scan, the training set of 3D models of 16 cadaveric specimens was used, and the weight of each variation mode was calculated based on the algorithm described earlier in "3D Model Reconstruction from DXA Scan" section. Next, to estimate the fracture risk for each subject ('test group'), 25-fold cross-validation was used (to allow maximum number of subjects in the training set as well as having equal number of fractured and non-fractured subjects in all groups). The 150 subjects were randomly divided into 25 groups. Each group consisted of two fractured cases and four non-fractured subjects. To predict the fracture risk for the subjects in each group (test group) the other 144 subjects ('training set group') were used to create and train a fracture risk prediction function (based on the reported fracture history of the subjects). This was done through a logistic regression analysis, which uses a logistic function to model a binary dependant variable (fracture vs. non-fracture). The variables used in the functions were the calculated modes, areal BMD, and the mean and standard deviation of pixels from the DXA scan. Subjects with an estimated probability of fracture greater than 50% (an arbitrary threshold) were considered high risk (likely to sustain a hip fracture).

Predicting the Fracture Risk based on 2D Model Reconstruction

Details regarding the 2D (i.e., DXA-based) statistical shape and appearance modeling have been described previously.[15] Briefly, landmarks were assigned to each of the DXA scans and then aligned and averaged to create the geometry template model. Next, each image was warped to the geometry template

TABLE 1. The summary of patients' characteristics.

Subjects	Total number	Male	Female	Age (mean $\pm$ SD)
Fractured	50	13	37	78.3 $\pm$ 7.4
Non-fractured	100	57	43	66.2 $\pm$ 9.5

model and the gray value of each pixel (which is an indication of the areal BMD value) was captured and normalized to the mean and standard deviation of all pixel values (within the same scan). All captured and normalized pixel values within the training set were then averaged to create the template BMD model. Principal Component Analysis (PCA) was used on both models (geometry and BMD) to find the main modes of variation for each and then combined, then PCA was again used to find the main modes of variation in describing the geometry and BMD distribution together. To reconstruct the geometry and BMD distribution of each DXA scan based on the variations in the training set, a series of landmarks on the contour of the femur were assigned to each DXA scan.[15] Then, the template geometry and BMD models were adjusted by the main modes of variation to recreate the DXA scan.[15] To estimate the fracture risk based on the 2D model reconstruction, the 25-fold cross-validation technique was conducted on the clinical data, as was done on the 3D model reconstructions.

Evaluation of the Fracture Risk Predictions

The two new image analysis methods (2D and 3D) were compared to two clinical metrics: total areal BMD and T-score. The total areal BMD from the DXA scans were also investigated using logistic regression analysis and 25-fold cross-validation in the same way as 2D and 3D SSAM. A threshold of 50% was used to assign each subject to high or low fracture risk. A T-score of -2.5 (the standard threshold for osteoporosis [4]) was also used to divide the subjects into low and high fracture risk groups. In the end, all predictions from 2D SSAM, 3D SSAM, BMD and T-score were compared to the fracture history of the subjects. To compare the 2D and 3D predictions, the average correct predictions were compared using a student t-test at a significance level of $\alpha = 0.05$.

To check the diagnostic value of each technique, the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (sensitivity) versus the false positive rate (1-specificity) based on different thresholds, was plotted and the area under the curve was calculated. To compare the geometry between the average fractured and non-fractured subjects, the mean location of each vertex was calculated for each group. The same was done for the BMD and to graphically illustrate the differences, colored heat maps were created for both.

RESULTS

Using a computer with a Core i7 processor and 16 GB RAM, the 2D analysis took less than one minute and 3D analysis took around 2 h to be completed for each scan. To account for more than 95% of the variation in describing the shape and BMD distribution of the cadaveric femurs nine and 14 modes were needed for 2D and 3D models, respectively. The average point to surface errors in the reconstruction of geometry was 1.65 ± 0.58 mm (range between 0.56 and 4.22 mm, Fig. 3), and the maximum error was related to the reconstruction of the greater trochanter. To depict the error proportionally to the geometry of the femur, it was normalized to the average widest anterior-posterior distance of the femurs in the training set (53 mm).

The average BMD reconstruction error for corresponding voxels (1 x 1 x 1 mm) was 0.11 ± 0.09 g/cm^3 (range between 0 and 0.84 g/cm^3), with the maximum error found in the cortical bone in the medial trochanteric area. The average BMD value from the 3D model reconstruction and the CT scans were illustrated for the mid-frontal plane and mid-transverse plane (Fig. 4).

In the clinical dataset, 2D SSAM was able to correctly classify 37 (out of 50) fractured cases and 92 (out of 100) non-fractured cases. Using 3D SSAM, the technique was able to correctly classify 38 (out of 50) fractured cases and 93 (out of 100) non-fractured cases. The T-score was able to correctly classify 18 (out of 50) fractured cases and 99 (out of 100) non-fractured cases, and the BMD correctly classified 34 (out of 50) fractured case and 93 (out of 100) non-fractured cases.

The average correct fracture risk prediction rate based on the 2D analysis for the fractured and non-fractured subjects were 0.74 ± 0.30, and 0.92 ± 0.22, respectively (Table 2). For the 3D analysis, these values were 0.76 ± 0.32, and 0.93 ± 0.26, and there were no statistically significant differences between the pre-
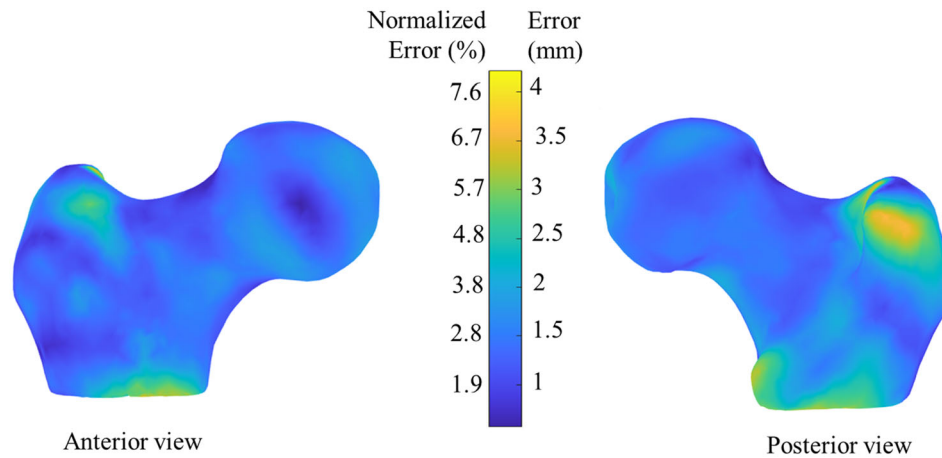
FIGURE 3. Illustration of the error in reconstruction of the geometry. The errors have been normalized to the average of the maximum thickness of the femurs in the training set to be able to compare it to the geometry of the femur as well. The maximum error was found at the tip of the greater trochanter.
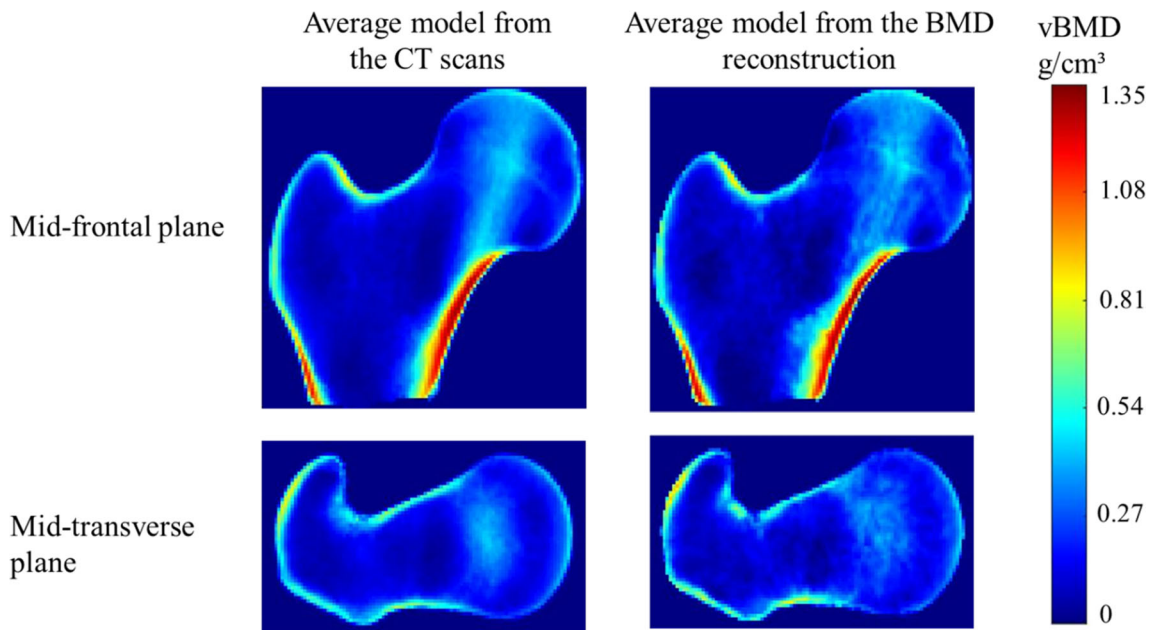

FIGURE 4. Illustration of the volumetric BMD (vBMD) in the average model from the CT scans and the average model from the BMD reconstruction in two views; top: mid-frontal plane, bottom: mid-transverse plane.

dictions based on 2D and 3D analysis for the fractured and non-fractured groups ($p = 0.83$ and $p = 0.76$). Also, both 2D and 3D analyses were able to improve hip fracture risk prediction for subjects who were at high risk of sustaining a hip fracture ($p < 0.0002$ and $p < 0.0001$, respectively).

The areas under the ROC curve for 2D SSAM, 3D SSAM, BMD, and T-score were calculated as 0.92, 0.91, 0.88, and 0.89 respectively, with 2D SSAM having the highest value and BMD having the lowest (Fig. 5). However, the pairwise comparison between the ROC curves didn't show a significant difference

between the areas under the curve for any two variables ($0.12 < p < 0.74$).

The differences between the average 3D shape and BMD distribution model for the fractured and non-fractured subjects were depicted using colored heat maps. In illustration of the differences in the geometry, if the average coordinates of the vertices in the non-fractured group were inside the average fractured geometry the distance was considered positive (i.e., non-fractured was smaller), and vice versa (Fig. 6). Generally, the average proximal femur's geometry for the fractured subjects had a greater outer diameter than the non-fractured one by an average of 0.6 mm.

BMES BIOMEDICAL ENGINEERING SOCIETY

TABLE 2. A summary of the hip fracture risk prediction results for various classifiers (mean±standard deviation).

Predictor	The rate of correct prediction for Fx subjects (%)	The rate of correct prediction for non-Fx subjects (%)	The rate of fracture risk under-prediction in all subjects (%)	The rate of fracture risk over-prediction in all subjects (%)
2D SSAM	74 ± 33	92 ± 12	9 ± 11	5 ± 8
3D SSAM	76 ± 33	93 ± 11	8 ± 11	5 ± 8
BMD	68 ± 37	93 ± 11	11 ± 13	5 ± 8
T-score	36 ± 38	99 ± 5	21 ± 11	1 ± 3

Fracture risk underprediction refers to the subjects that were identified as low risk but sustained a hip fracture and fracture risk overprediction refers to the subjects that were identified as high risk but didn't sustain a hip fracture.
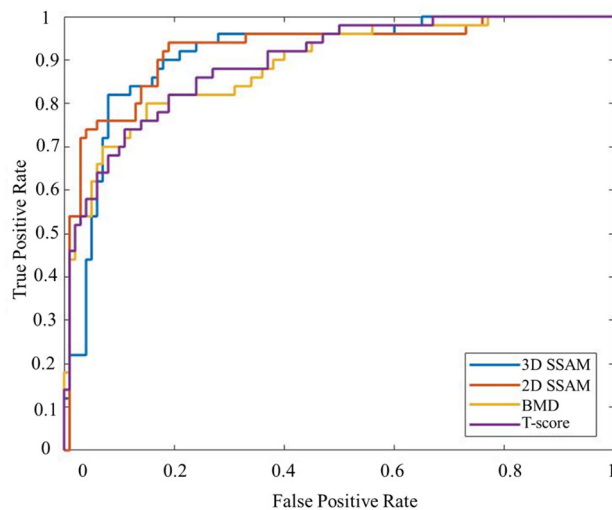*Fx* fractured, *non-Fx* non-fractured.

FIGURE 5. Receiver Operating Characteristic (ROC) curves for various hip fracture risk predictors. The area under the curve for 3D and 2D SSAM was slightly higher than BMD and T-score.

However, there was no statistically significant difference between the size of the models ($p = 0.07$).

For the BMD distribution comparison between the two groups (fractured and non-fractured subjects), the difference between the volumetric BMD of the voxels in the mid-frontal plane was calculated and depicted as a heat map as well, with higher BMD in the non-fractured group having a positive value. (Fig. 7). The average volumetric BMD map in the mid-frontal plane for the fractured subjects was lower than the non-fractured group, especially in the inner cortex of the trochanteric and subtrochanteric areas.

DISCUSSION

In this research, a novel approach to create a 3D model of the proximal femur from a single 2D DXA scan was introduced, evaluated, and its ability to
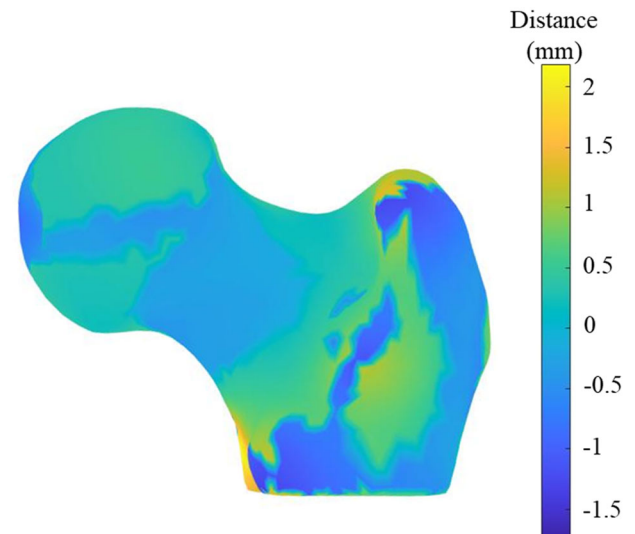

FIGURE 6. Surface geometry variation between the mean fractured and non-fractured subjects. The yellow color represents the points where the mean vertices of the non-fracture subjects were inside the mean fractured geometry (i.e., fractured group was larger than non-fractured) and the blue points indicate that the mean vertices of the non-fracture subjects were outside the mean fractured geometry (i.e., the mean fractured geometry was smaller than the non-fractured).

clinically predict hip fracture risk was assessed in a dataset of patients who were followed for at least five years. The new technique was able to significantly enhance hip fracture prediction in the high risk patients compared to T-score (40% improvement), which means that for the approximately 30,000 hip fractures that happen each year in Canada,[18] thousands of patients at high risk could be identified and protected from this injury by using this technique. In our previous studies, we have implemented 2D statistical modeling in a cadaveric study[15] and a clinical population study,[13] and the results showed that applying statistical models can greatly enhance hip fracture risk prediction in patients. In this study we showed that there was no real benefit to adding the 3D reconstruction for injury
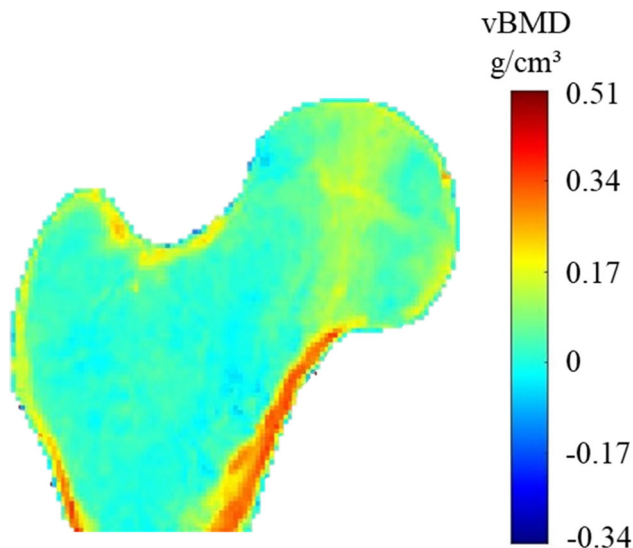
FIGURE 7. Volumetric BMD variation in the mid-frontal plane between the mean fractured and non-fractured subjects. The red color indicates the voxels that have a higher BMD value in the non-fractured subjects than the fractured subjects, and the blue color demonstrates the voxels that have a higher BMD value in the fractured subjects than the fractured ones.

risk prediction applications, making this easier and faster for clinical implementation. Also, this is the first known study to directly compare 2D vs 3D statistical shape and appearance modeling to predict the hip fracture risk in older adults. This has great importance since implementing 2D geometry and BMD distribution model reconstruction is associated with less computational burden and can be more achievable in clinical practice. These results can shape the future of applying statistical models in clinical practice to predict hip fracture risk.

Two previous studies have reported reconstruction errors in geometry and BMD distribution of similar magnitudes to those in the present study[35,38] (average geometry error of 1.07–1.1 mm, and an average BMD distribution error of 0.07–0.21 g/cm^3). However, the maximum geometry errors in this study were smaller than those previously reported (5.4–9.2 mm previous, vs. 4.2 mm herein).

Comparing the geometry of the proximal femur in the fractured and non-fractured subjects revealed that fractured cases tended to have a bigger outer diameter than non-fractured ones (Fig. 6), although this difference was not significant ($p = 0.07$). Other studies that have investigated the effect of the proximal femur's geometry on hip fracture risk[5,16,19,24,26] have found that there was an increase in the outer diameter of the femur in the fractured group. This effect could be attributed to the body's response to a decreased BMD and an effort to resist bending failure by increasing the diameter to increase the second moment of area.[2] It is

worth noting that that the range of the differences between the fractured and non-fractured geometries was between -1.5 and $+2$ mm, and considering that the average error in the geometry reconstruction was 1.6 mm, some of the difference between the two geometries might have been affected by the inherent error in the reconstruction. Therefore, the error in the geometry reconstruction in addition to the high coefficient of variation (CoV = 0.52) in the distance between the vertices in the two geometries (fractured and non-fractured) might have contributed to the lack of a statistically significant difference in the size of the femurs in the two groups in this study.

The average voxels' BMD in the mid-frontal plane in the fractured cases were lower than the non-fractured ones (Fig. 6). This could be specifically observed in the inner contour of the cortical bone in the medial region of the trochanteric area, which can be attributed to the thinning of the cortical bone in patients with osteoporosis.[23]

The area under the ROC curve for both 2D and 3D analyses were slightly better than T-score and BMD, although not statistically significant. When looking at the ROC curve, it can be observed that in the area of high specificity between 50 and 95% (close to the left side of the graph, 5–50% false positive rate) the statistical models were noticeably able to identify more true positive cases (people actually at risk of fracture) than the standard clinical method, which would be more desirable. It also showed that, only in the area of more than 50% false positive rate (close to the right side of the graph), the performance of all the methods were similar, and even in that case the T-score threshold should be modified from the -2.5 that is currently used. In practice, choosing the right threshold should be a trade-off between sensitivity and specificity and considering the cost of missing individuals at high risk or over treatment of people at low risk.

The area under the ROC curve in another similar 3D study[3] was reported as 0.83 for aBMD plus age, and 0.93 for 3D reconstruction (considering both geometry and BMD distribution) plus aBMD and age. However, two other studies that investigated the 2D analysis have reported 0.16,[1] and 0.03[6] improvement in area under the ROC curve while considering only the geometry, and geometry plus BMD distribution, respectively. These results suggest that comparing the improvement made by each method should be assessed based on various aspects of its performance, and for evaluation of different techniques a direct comparison based on the same training set and test set is preferred.

There could be several reasons for the lack of difference between 2D and 3D predictions. The most important one is that it might be possible that there is a

correlation between 2D and 3D geometry and BMD distribution features of the proximal femur. In some studies to reconstruct the 3D geometry of the proximal femur,[28,33] the main assumption was based on the dependency of 3D features on the 2D ones observed in a 2D image (either DXA scan or other radiographs of the hip), and their results showed that the 3D shape reconstruction of the proximal femur with this assumption had an acceptable average error range. Also, in this study some of the geometry traits were measured in 2D as well as 3D, and correlations of R^2 = 0.79, R^2 = 0.75, R^2 = 0.47, R^2 = 0.45 were found between the measurements in two directions of medial-lateral and anterior-posterior for the regions of (a) femoral head diameter, (b) neck diameter, (c) trochanter diameter 2 mm above lesser trochanter, and (d) trochanter diameter at the level of the lesser trochanter. So it could be concluded that the most of the 3D features of the proximal femur correspond with its 2D features, and although to describe a shape in 3D more variables are needed, most of these variables are correlated to ones observed in the 2D image.

One of the limitations of this study was that in the training set, the DXA scans and CT scans of isolated cadaveric femurs were used to make 2D and 3D template models, while for the evaluation of these techniques clinical DXA scans were used. The main difference between the clinical DXA scans and the ones from the isolated femurs was the effect of the overlapping pelvis over the proximal part of the femoral head, which led to artificially increasing the BMD measure in this area. Also, due to the presence of soft tissues in the clinical DXA scans, they were associated with more noise artifacts. Therefore, since these variabilities weren't captured in the training set, extra error might have been induced in the BMD distribution reconstruction model. However, the effect of these errors was minimized by using the clinical scans in creating the fracture risk estimation function through cross-validation.

Another limitation of this study was the limited number of femurs in the training set. This could limit the inclusion of all the geometric and BMD distribution traits of the proximal femur in the training set. Moreover, the predictions based on the statistical methods are heavily dependant on features and behaviors observed in the training set. It is therefore important that the training set include many of the characteristics that are present in a population with respect to sex, ethnicity, and age. However, in this study with the limited number of the scans in the training set, the fracture risk prediction was improved, and one can expect that with a more comprehensive training set the fracture risk estimation could be even enhanced more.

Also, in order to be able to implement any complementary technique in clinical practice to enhance the estimation of the hip fracture risk, first it should be validated in various studies, and next, it should be offered in a user-friendly platform with minimum user-interference (to minimize the user-induced variability). Therefore, further studies based on comprehensive training sets are still required to support implementing statistical models in clinical practice. These methods should also be presented in platforms that are compatible with the current techniques used in practice.

In addition to the mechanical properties of the proximal femur, many other factors affect a patients' hip fracture risk. These factors either relate to the patients' characteristics[34] (e.g., medication use, fracture history, tobacco use, alcohol consumption), fall mechanics[21] (e.g., patients' height, weight, and reflexes), or fall probability[30] (e.g., physical activity level, comorbidities, balance and stability, and age). However, in this research, only features related to the proximal femurs' structural strength were investigated. Therefore, a more robust prediction would consider many of these other factors.

This study showed that, while proximal femurs 3D model reconstruction might be necessary for further numerical analysis (e.g., finite element analysis and direct measurement of specific 3D traits), it doesn't add significant value to the hip fracture risk estimation when compared to 2D model reconstruction. This will have a significant impact on how statistical models are adopted by clinical practice. Since implementing 2D techniques is less intensive technically and computationally, and uses more accessible and safer imaging modalities (compared to using CT scans) to expand the training set, it has great potential to be implemented in clinical practice as part of standard hip fracture risk estimation in older adults.

REFERENCES

[1]Baker-Lepain, J. C., K. R. Luker, J. A. Lynch, N. Parimi, M. C. Nevitt, and N. E. Lane. Active shape modeling of the hip in the prediction of incident hip fracture. *J. Bone Miner. Res.* 2011. https://doi.org/10.1002/jbmr.254.

[2]Bono, C. M., and T. A. Einhorn. Overview of osteoporosis: pathophysiology and determinants of bone strength. *Aging Spine* 2003. https://doi.org/10.1007/s00586-003-0603-2.

[3]Bredbenner, T. L., R. L. Mason, L. M. Havill, E. S. Orwoll, and D. P. Nicolella. Fracture risk predictions based on statistical shape and density modeling of the proximal femur. *J. Bone Miner. Res.* 2014. https://doi.org/10.1002/jbmr.2241.

[4]Cosman, F., S. J. de Beur, M. S. LeBoff, E. M. Lewiecki, B. Tanner, S. Randall, and R. Lindsay. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos. Int.* 2014. https://doi.org/10.1007/s00198-014-2794-2.

[5]Gnudi, S., C. Ripamonti, G. Gualtieri, and N. Malavolta. Geometry of proximal femur in the prediction of hip fracture in osteoporotic women. *Br. J. Radiol.* 1999. https://doi.org/10.1259/bjr.72.860.10624337.

[6]Goodyear, S. R., R. J. Barr, E. McCloskey, S. Alesci, R. M. Aspden, D. M. Reid, and J. S. Gregory. Can we improve the prediction of hip fracture by assessing bone structure using shape and appearance modelling? *Bone* 53:188–193, 2013.

[7]Grassi, L., S. P. Väänänen, M. Ristinmaa, J. S. Jurvelin, and H. Isaksson. Prediction of femoral strength using 3D finite element models reconstructed from DXA images: validation against experiments. *Biomech. Model. Mechanobiol.* 16:989–1000, 2017.

[8]Gregory, J. S., and R. M. Aspden. Femoral geometry as a risk factor for osteoporotic hip fracture in men and women. *Med. Eng. Phys.* 2008. https://doi.org/10.1016/j.medengphy.2008.09.002.

[9]Hadjimichael, A. C. Hip fractures in the elderly without osteoporosis. *J. Frailty* 3:8–12, 2018.

[10]Humbert, L., T. Whitmarsh, M. De Craene, L. M. Del Río Barquero, K. Fritscher, R. Schubert, F. Eckstein, T. Link, and A. F. Frangi. 3D reconstruction of both shape and bone mineral density distribution of the femur from DXA images. *2010 7th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2010 - Proc.* 456–459, 2010. https://doi.org/10.1109/isbi.2010.5490310.

[11]Humbert, L., R. Winzenrieth, S. Di Gregorio, T. Thomas, L. Vico, J. Malouf, and L. M. del Río Barquero. 3D analysis of cortical and trabecular bone from hip DXA: precision and trend assessment interval in postmenopausal women. *J. Clin. Densitom.* 2019. https://doi.org/10.1016/j.jocd.2018.05.001.

[12]Ip, T. P., J. Leung, and A. W. C. Kung. Management of osteoporosis in patients hospitalized for hip fractures. *Osteoporos. Int.* 2010. https://doi.org/10.1007/s00198-010-1398-8.

[13]Jazinizadeh, F., J. D. Adachi, and C. E. Quenneville. Advanced 2D image processing technique to predict hip fracture risk in an older population based on single DXA scans. *Osteoporos. Int.* 31:1925–1933, 2020. https://doi.org/10.1007/s00198-020-05444-7.

[14]Jazinizadeh, F., H. Mohammadi, and C. E. Quenneville. Comparing the fracture limits of the proximal femur under impact and quasi-static conditions in simulation of a sideways fall. *J. Mech. Behav. Biomed. Mater.* 103:103593, 2020.

[15]Jazinizadeh, F., and C. E. Quenneville. Enhancing hip fracture risk prediction by statistical modeling and texture analysis on DXA images. *Med. Eng. Phys.* 78:14–20, 2020. https://doi.org/10.1016/j.medengphy.2020.01.015.

[16]Kaptoge, S., T. J. Beck, J. Reeve, K. L. Stone, T. A. Hillier, J. A. Cauley, and S. R. Cummings. Prediction of incident hip fracture risk by femur geometry variables measured by hip structural analysis in the study of osteoporotic fractures. *J. Bone Miner. Res.* 2008. https://doi.org/10.1359/jbmr.080802.

[17]Le Corroller, T., M. Pithioux, F. Chaari, B. Rosa, S. Parratte, B. Maurel, J. N. Argenson, P. Champsaur, and P. Chabrand. Bone texture analysis is correlated with three-dimensional microarchitecture and mechanical properties of trabecular bone in osteoporotic femurs. *J. Bone Miner. Metab.* 2013. https://doi.org/10.1007/s00774-012-0375-z.

[18]Leslie, W. D., S. O'Donnell, C. Lagacé, P. Walsh, C. Bancej, S. Jean, K. Siminoski, S. Kaiser, D. L. Kendler, and S. Jaglal. Population-based Canadian hip fracture rates with international comparisons. *Osteoporos. Int.* 2010. https://doi.org/10.1007/s00198-009-1080-1.

[19]Michelotti, J., and J. M. Clark. Femoral neck length and hip fracture risk. *J. Bone Miner. Res.* 1999. https://doi.org/10.1359/jbmr.1999.14.10.1714.

[20]Miller, P. D., S. Barlas, S. K. Brenneman, T. A. Abbott, Y. T. Chen, E. Barrett-Connor, and E. S. Siris. An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch. Int. Med.* 2004. https://doi.org/10.1001/archinte.164.10.1113.

[21]Nankaku, M., H. Kanzaki, T. Tsuboyama, and T. Nakamura. Evaluation of hip fracture risk in relation to fall direction. *Osteoporos. Int.* 16:1315–1320, 2005.

[22]Nasiri Sarvi, M., and Y. Luo. Sideways fall-induced impact force and its effect on hip fracture risk: a review. *Osteoporos. Int.* 28:2759–2780, 2017.

[23]Nguyen, B. N. T., H. Hoshino, D. Togawa, and Y. Matsuyama. Cortical thickness index of the proximal femur: a radiographic parameter for preliminary assessment of bone mineral density and osteoporosis status in the age 50 years and over population. *CiOS Clin. Orthop. Surg.* 2018. https://doi.org/10.4055/cios.2018.10.2.149.

[24]Osterhoff, G., E. F. Morgan, S. J. Shefelbine, L. Karim, L. M. McNamara, and P. Augat. Bone mechanical properties and changes with osteoporosis. *Injury* 2016. https://doi.org/10.1016/s0020-1383(16)47003-8.

[25]Parker, M. J., W. J. Gillespie, and L. D. Gillespie. Effectiveness of hip protectors for preventing hip fractures in elderly people: systematic review. *Bmj* 2006. https://doi.org/10.1136/bmj.38753.375324.7c.

[26]Ripamonti, C., L. Lisi, and M. Avella. Femoral neck shaft angle width is associated with hip-fracture risk in males but not independently of femoral neck bone density. *Br. J. Radiol.* 2014. https://doi.org/10.1259/bjr.20130358.

[27]Roche, J. J. W., R. T. Wenn, O. Sahota, and C. G. Moran. Effect of comorbidities and postoperative complications on mortality after hip fracture in elderly people: prospective observational cohort study. *Br. Med. J.* 2005. https://doi.org/10.1136/bmj.38643.663843.55.

[28]Schumann, S., M. Tannast, L. P. Nolte, and G. Zheng. Validation of statistical shape model based reconstruction of the proximal femur—a morphology study. *Med. Eng. Phys.* 2010. https://doi.org/10.1016/j.medengphy.2010.03.010.

[29]Siris, E. S., Y. T. Chen, T. A. Abbott, E. Barrett-Connor, P. D. Miller, L. E. Wehren, and M. L. Berger. Bone min-

eral density thresholds for pharmacological intervention to prevent fractures. *Arch. Intern. Med.* 2004. https://doi.org/10.1001/archinte.164.10.1108.

[30]Slemenda, C., S. Cummings, E. Seeman, P. Lips, D. Black, and D. B. Karpf. Prevention of hip fractures: risk factor modification. *Am. J. Med.* 1997. https://doi.org/10.1016/s0002-9343(97)90028-0.

[31]Sozen, T., L. Ozisik, and N. Calik Basaran. An overview and management of osteoporosis. *Eur. J. Rheumatol.* 2017. https://doi.org/10.5152/eurjrheum.2016.048.

[32]Thevenot, J., J. Hirvasniemi, P. Pulkkinen, M. Määttä, R. Korpelainen, S. Saarakkala, and T. Jämsä. Assessment of risk of femoral neck fracture with radiographic texture parameters: a retrospective study. *Radiology* 2014. https://doi.org/10.1148/radiol.14131390.

[33]Thevenot, J. Ô., J. Koivumäki, V. Kuhn, F. Eckstein, and T. Jämsä. A novel methodology for generating 3D finite element models of the hip from 2D radiographs. *J. Biomech.* 47:438–444, 2014.

[34]Unnanuntana, A., B. P. Gladnick, E. Donnelly, and J. M. Lane. The assessment of fracture risk. *j. Bone Joint Surg.* 2010. https://doi.org/10.2106/jbjs.i.00919.

[35]Väänänen, S. P., L. Grassi, G. Flivik, J. S. Jurvelin, and H. Isaksson. Generation of 3D shape, density, cortical thickness and finite element mesh of proximal femur from a DXA image. *Med. Image Anal.* 24:125–134, 2015.

[36]Wainwright, S. A., L. M. Marshall, K. E. Ensrud, J. A. Cauley, D. M. Black, T. A. Hillier, M. C. Hochberg, M. T. Vogt, and E. S. Orwoll. Hip fracture in women without osteoporosis. *J. Clin. Endocrinol. Metab.* 2005. https://doi.org/10.1210/jc.2004-1568.

[37]Whitmarsh, T., K. D. Fritscher, L. Humbert, L. M. del Río Barquero, T. Roth, C. Kammerlander, M. Blauth, R. Schubert, and A. F. Frangi. Hip fracture discrimination from dual-energy X-ray absorptiometry by statistical model registration. *Bone* 51:896–901, 2012.

[38]Whitmarsh, T., L. Humbert, M. De Craene, L. M. del Río Barquero, and A. F. Frangi. Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy x-ray absorptiometry. *IEEE Trans. Med. Imaging* 30:2101–2114, 2011.

[39]Woolf, A. D., and K. Åkesson. Preventing fractures in elderly people. *bmj* 30:2101–2114, 2003.

[40]Yang, L., A. C. Burton, M. Bradburn, C. M. Nielson, E. S. Orwoll, and R. Eastell. Distribution of bone density in the proximal femur and its association with hip fracture risk in older men: the osteoporotic fractures in men (MrOS) study. *J. Bone Miner. Res.* 2012. https://doi.org/10.1002/jbmr.1693.

[41]Yang, L., L. Palermo, D. M. Black, and R. Eastell. Prediction of incident hip fracture with the estimated femoral strength by finite element analysis of DXA scans in the study of osteoporotic fractures. *J. Bone Miner. Res.* 29:2594–2600, 2014.

[42]Yang, L., W. J. M. Udall, E. V. McCloskey, and R. Eastell. Distribution of bone density and cortical thickness in the proximal femur and their association with hip fracture in postmenopausal women: a quantitative computed tomography study. *Osteoporos. Int.* 2014. https://doi.org/10.1007/s00198-013-2401-y.