

# Women with previous fragility fractures can be classified based on bone microarchitecture and finite element analysis measured with HR-pQCT

K. K. Nishiyama · H. M. Macdonald · D. A. Hanley · S. K. Boyd

Received: 15 July 2012 / Accepted: 13 September 2012 / Published online: 20 November 2012  
© International Osteoporosis Foundation and National Osteoporosis Foundation 2012

## Abstract

**Summary** High-resolution peripheral quantitative computed tomography (HR-pQCT) measurements of distal radius and tibia bone microarchitecture and finite element (FE) estimates of bone strength performed well at classifying postmenopausal women with and without previous fracture. The HR-pQCT measurements outperformed dual energy x-ray absorptiometry (DXA) at classifying forearm fractures and fractures at other skeletal sites.

**Introduction** Areal bone mineral density (aBMD) is the primary measurement used to assess osteoporosis and fracture risk; however, it does not take into account bone microarchitecture, which also contributes to bone strength. Thus, our objective was to determine if bone microarchitecture measured with HR-pQCT and FE estimates of bone strength could classify women with and without low-trauma fractures.

**Methods** We used HR-pQCT to assess bone microarchitecture at the distal radius and tibia in 44 postmenopausal women with a history of low-trauma fracture and 88 age-matched controls from the Calgary cohort of the Canadian Multicentre Osteoporosis Study (CaMos) study. We estimated bone strength using FE analysis and simulated distal radius aBMD from the HR-pQCT scans. Femoral neck (FN) and lumbar spine (LS) aBMD were measured with DXA. We used support vector machines (SVM) and a tenfold cross-validation to classify the fracture cases and controls and to determine accuracy.

**Results** The combination of HR-pQCT measures of microarchitecture and FE estimates of bone strength had the highest area under the receiver operating characteristic (ROC) curve of 0.82 when classifying forearm fractures compared to an area under the curve (AUC) of 0.71 from DXA-

---

K. K. Nishiyama · S. K. Boyd  
Schulich School of Engineering, University of Calgary,  
Calgary, Canada

K. K. Nishiyama  
e-mail: kksnishi@ucalgary.ca

K. K. Nishiyama · S. K. Boyd  
Roger Jackson Centre for Health and Wellness Research,  
University of Calgary,  
Calgary, Canada

K. K. Nishiyama · D. A. Hanley · S. K. Boyd  
McCaig Institute for Bone and Joint Health,  
Calgary, Canada

D. A. Hanley  
e-mail: dahanley@ucalgary.ca

H. M. Macdonald  
Department of Orthopedics, University of British Columbia,  
Vancouver, Canada  
e-mail: heather.macdonald@ubc.ca

H. M. Macdonald  
Child and Family Research Institute,  
Vancouver, Canada

H. M. Macdonald  
Centre for Hip Health and Mobility,  
Vancouver, Canada

D. A. Hanley · S. K. Boyd  
Faculty of Medicine, University of Calgary,  
Calgary, Canada

S. K. Boyd (✉)  
Department of Mechanical and Manufacturing Engineering,  
University of Calgary,  
2500 University Drive,  
N.W. Calgary, Alberta, Canada T2N 1N4  
e-mail: skboyd@ucalgary.ca

derived aBMD of the forearm and 0.63 from FN and spine DXA. For all fracture types, FE estimates of bone strength at the forearm alone resulted in an AUC of 0.69.

**Conclusion** Models based on HR-pQCT measurements of bone microarchitecture and estimates of bone strength performed better than DXA-derived aBMD at classifying women with and without prior fracture. In future, these models may improve prediction of individuals at risk of low-trauma fracture.

**KEYWORDS** Bone microarchitecture · Finite element analysis · Fracture · HR-pQCT · Support vector machines

## Introduction

Areal bone mineral density (aBMD) measured by dual energy x-ray absorptiometry (DXA) is a significant predictor of fracture risk [1]; however, half of the fractures occur in women who would not be classified as osteoporotic by aBMD [2]. Mechanical strength of bone is dependent on bone matrix, apparent density, and microarchitecture [3] and measurements of cortical and trabecular microarchitecture are independent of aBMD [4]. Thus, incorporating bone structural information with density measurements could potentially improve assessment of fracture risk.

High-resolution peripheral quantitative computed tomography (HR-pQCT) is an emerging technology capable of measuring cortical and trabecular microarchitecture at distal radius and distal tibia [5, 6]. The resolution of HR-pQCT also permits application of finite element (FE) analysis to the 3D scans to estimate bone strength. These FE models are highly correlated with bone strength measured directly with mechanical testing [7] as they incorporate both material and architectural components of bone strength. Previous HR-pQCT studies examined the ability of individual bone microarchitectural [8–11] and FE analysis parameters [8–10] or a combination of parameters [9, 10] to discriminate between postmenopausal women with and without a history of low-trauma fracture. These studies showed that bone structure and strength contribute to forearm fracture risk independently of aBMD [8]. It has been an important first step to establish the association of microarchitectural parameters and FE outcomes with low-trauma fractures [10]. This naturally leads us to determine whether it is possible to classify people with low-trauma fracture based on similar outcomes.

Machine learning methods are statistical tools used to recognize patterns in datasets. These methods are often used for the purpose of classification since they can be trained based on known cases and then tested on new cases. Machine learning methods can be used to incorporate all HR-pQCT and FE parameters in order to classify those with and without fracture. This is in contrast to statistical approaches, such as

principal component analysis (PCA), which reduce data into combinations of parameters that account for majority of variance. Atkinson et al. recently used a machine learning method, gradient boosting machines (GBM), and found that fracture prediction improved when all possible bone density, geometry, and microstructural parameters obtained with central quantitative computed tomography (QCT) and HR-pQCT were included in the GBM model when compared with DXA parameters alone [12]. Support vector machines (SVM) are an alternative machine learning method that can be used to classify individuals with and without previous fracture by maximizing the separation between groups. SVM models offer several advantages over other machine learning methods, including greater stability and minimal requirements for parameter tuning. Thus, our objective was to use SVM models to incorporate bone microarchitectural parameters and FE estimates of bone strength to determine if we can classify women with and without previous low trauma fractures.

## Materials and methods

### Participants

Participants in this study were postmenopausal women who were members of the Calgary, AB cohort of the Canadian Multicentre Osteoporosis Study (CaMos) [13, 14]. Briefly, CaMos is a 10-year prospective population-based study in which participants were recruited using a stratified random sampling technique from nine centers across Canada [15]. At the 10-year follow-up, we invited individuals from the Calgary CaMos cohort to participate in an HR-pQCT sub-study. From this cohort ( $N=442$ ; ages 20–99 years), we identified postmenopausal women who sustained a low-trauma fracture (excluding finger, face, and toes) during the 10 years of study follow-up ( $n=44$ ). Fractures were identified by self-report during scheduled interviews at years 3, 5, and 10 and by yearly postal questionnaires through year 9. Additional information regarding circumstances surrounding the fracture and fracture site were gathered via a structured telephone interview [16]. Low-trauma fractures were those that occurred without trauma or as a result of a fall from standing height or less. We randomly matched two participants without fracture by age ( $\pm 1.0$  year) to each fracture case ( $n=88$ ). We determined height, weight, bisphosphonate and corticosteroid use ( $>3$  months), and menopause status from the interviewer-administered CaMos questionnaire. In addition, as part of the standard CaMos assessment, trained technicians measured femoral neck (FN) and lumbar spine (LS) aBMD (gram per square centimeter) using DXA (Hologic QDR4500, Bedford, USA). The Conjoint Health Research Ethics Board at the University of Calgary approved all procedures.

## HR-pQCT scan acquisition and measurements

As we described in detail elsewhere [13, 14], we scanned all participants using HR-pQCT (XtremeCT; SCANCO Medical, Brüttisellen, Switzerland) at the nondominant radius and left tibia unless there was a previous fracture at the desired site, in which case we scanned the opposite limb. One of two highly trained operators acquired and analyzed all scans according to the manufacturer's standard in vivo protocol. We scanned all participants using 60 kVp effective energy, 1,000  $\mu$ A current, and 100 ms integration time to acquire 110 slices (approximately 9.02 mm) of the radius and tibia at an 82  $\mu$ m nominal isotropic resolution. Scans were manually scored for motion on a scale of 0 (no motion) to 4 (significant blurring of the periosteal surface, discontinuities in the cortical shell, or streaking in the soft tissue) and scans scored as a 4 were excluded from this analysis ( $n=2$ ).

We used the manufacturer's standard method to filter and binarize the HR-pQCT images [17] and assessed all standard HR-pQCT morphological microstructure outcomes [5]. These measurements were previously validated against microcomputed tomography [6, 18] and have in vivo short-term reproducibility of <4.5 % in adult populations [19]. Reproducibility in our lab is <3.8 % for all parameters at the radius and tibia.

To segment the cortical and trabecular regions, we used an automatic segmentation algorithm implemented in Image Processing Language (IPL V5.07, SCANCO Medical, Brüttisellen, Switzerland) [20]. Based on this segmentation, we calculated macrostructural parameters: cortical (Ct.Ar, square millimeter), trabecular (Tb.Ar, square millimeter), and total (Tt.Ar, square millimeter) cross-sectional areas. In addition, we calculated cortical porosity (Ct.Po, percent) as the number of void voxels within the cortex [14, 20] and measured direct cortical thickness (Ct.Th, millimeter), and cortical bone mineral density (Ct.BMD, milligram hydroxyapatite (HA) per cubic centimeter).

We acquired FN and LS aBMD by DXA but did not acquire ultradistal radius (UDR) DXA scans for the participants. Therefore, we implemented the method developed by Burghardt et al. [21] to simulate DXA-derived aBMD: we aligned the HR-pQCT scans in the same configuration as a DXA scan and obtained a calibrated projection image from which we determined aBMD. The aBMD values determined by HR-pQCT were then calibrated based on the relationship reported by Burghardt et al. [21]. This method is strongly correlated with DXA aBMD ( $R^2=0.82$ ) and is highly reproducible with a root mean-squared coefficient of variation of 1.1 %.

## Finite element analysis

From the HR-pQCT images, we generated FE meshes using the voxel conversion approach [22, 23]. Each voxel was converted to an eight-node, 82  $\mu$ m<sup>3</sup>, hexahedral element. The

meshes generated resulted in approximately 1 million elements for the radius scans and 2.5 million elements for the tibia scans. We simulated uniaxial compression on each radius and tibia model up to 1 % strain using a Young's modulus of 6,829 MPa and Poisson's ratio of 0.3 [7] as previously reported [13, 14]. We used a custom FE solver (FAIM, Version 4.0; Numerics88 Solutions, Calgary, Canada) on a desktop workstation (Mac Pro, OSX, Version 10.5.6; 2 $\times$ 2.8 GHz Quad-Core Intel Xenon) to estimate bone strength (ultimate stress, megapascal) based on the relationship determined by MacNeil and Boyd [7]. We also determined the percentage of load carried by the cortex based on the elements labeled as cortical bone from the automatic segmentation at the most distal and most proximal slices. In order to estimate the risk of forearm fracture, we calculated the load-to-strength ratio ( $\Phi$ ) [24–26]. This ratio represents the estimated fall force on the outstretched arm divided by the estimated failure load from the FE analysis. The theoretical fracture threshold occurs when  $\Phi > 1.0$  [24–26].

## Statistical analysis

We used SVM [27] with a weighted radial basis kernel function to classify women with and without low-trauma fracture. In the first analysis, we examined only low-trauma forearm fractures and in the second analysis, we included all low-trauma fractures. We used Student's *t* test to compare outcome variables between the fracture and control groups. For each analysis, we generated SVM models based on standardized measurements of (1) all HR-pQCT, FE, and DXA parameters; (2) all HR-pQCT and FE parameters; (3) FE parameters alone; (4) UDR, LS, and FN aBMD; and (5) LS and FN aBMD. For the forearm fractures, we used only the distal radius measurements and for all types of fracture we used the distal radius and distal tibia measurements. To gauge the accuracy of the SVM results and to avoid overfitting to our specific dataset, we used a tenfold cross-validation scheme. This method randomly divides the participants into ten groups, trains the model on nine of these groups, and tests the model on the remaining group. This is repeated ten times so that each group is used as the test group and the accuracy reported is the average of all iterations. The results of this validation are presented as the areas under the curve (AUC) for receiver operating characteristic (ROC) curves. Weka (version 3.7; University of Waikato, Hamilton, New Zealand) was used for the SVM analyses [28], and SPSS Statistics (Version 19.0; IBM; Somers, NY, USA) for all other analyses.

## Results

### Participant characteristics and bone outcome variables

Descriptive characteristics of the participants are provided in Table 1. The fracture and nonfracture groups were similar

with respect to age, height, weight, body mass index (BMI), and use of corticosteroids. Use of bisphosphonates was higher among women with a history of forearm fracture compared with nonfracture controls ( $p=0.028$ ). The median time between fracture and scan date was 4.5 years (interquartile range (IQR): 3.2–6.7 years). The majority of fractures were forearm fractures ( $n=14$ ) and lower leg ( $n=14$ ), followed by upper arm ( $n=9$ ), spine ( $n=5$ ), hip ( $n=4$ ), upper leg ( $n=4$ ), and ribs ( $n=2$ ). Eight women reported two low-trauma fractures; we classified these women based on their first fracture.

We present the HR-pQCT and FE variables in Table 2 and DXA parameters in Table 3. Significant differences ( $p<0.05$ ) between the forearm fracture group and nonfracture controls were found at the radius for all parameters with the exception of Tb.Th, Ct.Po, and Ct.Ar. When all types of fractures were pooled, all parameters were significantly different with the exception of Tb.Th, Ct.Po, Ct.BMD, Ct.Th, and area measurements. At the distal tibia, Tb.N, Tb.Sp, Ct.BMD, Ct.Po, Tt.Ar, Tb.Ar, and the percentage of the load carried by the cortical region at the distal and proximal slices were not significantly different between forearm fracture cases and controls. For all fracture types pooled at the distal tibia, Ct.BMD, Ct.Po, Tt.Ar, Tb.Ar, and the percentage of the load carried by the cortical region were not significantly different between fracture cases and controls.

The percentage of women who had load-to-strength ratios greater than 1.0 and, therefore, theoretically at greater risk of fracture was 28.6 % for those with forearm fractures and 10.7 % for their corresponding nonfracture controls; however, the mean of the entire groups were still both below 1.0 (Table 2). For women with all fracture types, 25.0 % exceeded a ratio of 1.0 while 13.6 % of their corresponding controls exceeded the threshold. Women with forearm fracture had significantly lower LS aBMD and UDR aBMD compared with the controls. The women with fractures at all sites had significantly lower FN and LS aBMD from their corresponding controls.

## SVM classification

The accuracy, sensitivity, specificity, and ROC-AUC's from the cross-validation of the SVM models are summarized in Table 4. For the forearm fractures and controls, the highest accuracy (83.3 %) and AUC (0.82) was achieved using the HR-pQCT measurements and FE estimates of bone strength together. This was followed by the combination of HR-pQCT, FE, and DXA parameters (accuracy: 81.0 %, AUC: 0.80) and by FE parameters alone (accuracy: 78.6 %, AUC: 0.80). Using LS, FN, and UDR aBMD produced an accuracy of 71.4 % and an AUC of 0.71, while using only LS and FN aBMD had the worst performance (accuracy: 59.5 %, AUC: 0.63) classifying the women. When considering low-trauma fractures at all sites, the SVM models were not as effective at classifying the fracture and control participants compared with the focused forearm fracture analysis. The FE parameters alone had the highest accuracy (69.7 %) and had the highest AUC (0.69). The combination of HR-pQCT and FE measurements had an accuracy of 68.9 % and an AUC of 0.67 while the HR-pQCT, FEA, and DXA aBMD also had an accuracy of 68.9 % and an AUC of 0.68.

## Discussion

In this study, we successfully demonstrated that a machine learning technique based on SVM model could classify postmenopausal women with and without a history of low-trauma fracture based on HR-pQCT input. The SVM model incorporating HR-pQCT bone microarchitecture and density, as well as FE estimates of bone strength at the distal radius, outperformed DXA-derived aBMD, which is the clinical gold standard for osteoporosis screening. The SVM method is a promising tool for classification of women with and without previous low-trauma fracture and could potentially be an important clinical basis for fracture risk assessment (Fig. 1).

SVMs are a type of machine learning tool that recognize patterns in data and map data to a higher dimension to create

**Table 1** Descriptive characteristics of postmenopausal women with forearm fractures and fractures at all sites, and their corresponding age-matched controls

	Control ( $n=28$ )	Forearm low-trauma fx ( $n=14$ )	$p$ value	Control ( $n=88$ )	Any low-trauma fx ( $n=44$ )	$p$ value
Age (SD) (years)	72.8 (8.5)	73.1 (8.5)	0.91	73.3 (7.8)	73.4 (7.5)	0.99
Height (SD) (cm)	161.2 (6.1)	161.6 (4.8)	0.82	158.5 (6.9)	158.7 (6.9)	0.88
Weight (SD) (kg)	72.2 (13.2)	71.7 (12.5)	0.90	70.5 (15.7)	70.1 (12.4)	0.88
BMI (SD) ( $\text{kg}/\text{m}^2$ )	27.8 (5.2)	27.4 (4.0)	0.77	28.1 (6.8)	27.8 (4.6)	0.77
Corticosteroid use ( $>3$ months) (yes/no, %)	5/23 (17.9)	2/12 (14.3)	0.78	13/75 (14.7)	4/40 (9.1)	0.36
Bisphosphonate use (yes/no, %)	12/16 (42.9)	11/3 (78.6)	0.03	42/46 (47.7)	15/29 (34.1)	0.14
Years since menopause (SD)	25.3 (12.0)	26.4 (11.2)	0.61	24.0 (13.0)	27.2 (13.3)	0.46

**Table 2** Bone microarchitecture and finite element parameters [mean (SD)] for the fracture and control groups at the distal radius and distal tibia

	Control (n=28)	Forearm low-trauma fx (n=14)	p value	Control (n=88)	Any low-trauma fx (n=44)	p value
<b>Distal radius</b>						
Tt.BMD (mg HA/cm <sup>3</sup> )	298 (67)	210 (49)	<0.001	283 (71)	243 (63)	0.002
Tb.BMD (mg HA/cm <sup>3</sup> )	152 (39)	94 (30)	<0.001	146 (41)	119 (44)	<0.001
BV/TV	0.127 (0.032)	0.078 (0.025)	<0.001	0.122 (0.034)	0.099 (0.036)	<0.001
Tb.N (mm <sup>-1</sup> )	1.83 (0.30)	1.24 (0.43)	<0.001	1.81 (0.36)	1.53 (0.47)	<0.001
Tb.Th (mm)	0.069 (0.012)	0.066 (0.018)	0.53	0.067 (0.012)	0.065 (0.015)	0.57
Tb.Sp (mm)	0.492 (0.100)	0.840 (0.317)	<0.001	0.520 (0.197)	0.677 (0.317)	<0.001
Ct.BMD (mg HA/cm <sup>3</sup> )	782 (87)	711 (96)	0.02	762 (92)	731 (78)	0.06
Ct.Th (mm)	0.99 (0.17)	0.84 (0.21)	0.02	0.94 (0.19)	0.87 (0.19)	0.06
Ct.Po (%)	10.2 (4.7)	11.5 (5.6)	0.42	11.2 (5.0)	11.7 (4.6)	0.57
Tt.Ar (mm <sup>2</sup> )	264.7 (47.5)	295.9 (42.8)	0.05	266.5 (48.0)	274.2 (49.2)	0.39
Ct.Ar (mm <sup>2</sup> )	58.9 (8.3)	55.0 (9.9)	0.18	56.7 (9.7)	54.9 (9.9)	0.30
Tb.Ar (mm <sup>2</sup> )	205.8 (44.9)	240.9 (43.8)	0.02	209.8 (47.4)	219.3 (48.8)	0.28
Ultimate stress (MPa)	24.1 (9.0)	13.0 (5.3)	<0.001	22.3 (8.6)	17.4 (8.3)	0.002
Load cortex distal (%)	40.8 (8.1)	46.5 (8.9)	0.05	41.6 (9.0)	46.0 (9.7)	0.01
Load cortex prox (%)	86.4 (8.3)	94.5 (5.8)	0.002	86.4 (8.6)	90.7 (7.6)	0.01
Load-to-strength ratio	0.73 (0.23)	0.99 (0.31)	0.01	0.77 (0.21)	0.90 (0.27)	0.004
<b>Distal tibia</b>						
Tt.BMD (mg HA/cm <sup>3</sup> )	268 (52)	211 (27)	<0.001	256 (59)	219 (49)	<0.001
Tb.BMD (mg HA/cm <sup>3</sup> )	163 (35)	130 (27)	0.003	159 (37)	136 (39)	0.001
BV/TV	0.136 (0.029)	0.109 (0.023)	0.003	0.132 (0.031)	0.115 (0.034)	0.004
Tb.N (mm <sup>-1</sup> )	1.73 (0.35)	1.58 (0.36)	0.22	1.75 (0.37)	1.54 (0.48)	0.01
Tb.Th (mm)	0.080 (0.016)	0.070 (0.011)	0.03	0.077 (0.015)	0.078 (0.020)	0.78
Tb.Sp (mm)	0.528 (0.150)	0.612 (0.244)	0.17	0.538 (0.240)	0.695 (0.460)	0.01
Ct.BMD (mg HA/cm <sup>3</sup> )	728 (82)	687 (61)	0.11	705 (115)	685 (68)	0.30
Ct.Th (mm)	1.47 (0.32)	1.22 (0.16)	0.01	1.39 (0.34)	1.25 (0.25)	0.01
Ct.Po (%)	16.6 (5.4)	17.2 (4.6)	0.71	17.5 (6.6)	18.0 (4.6)	0.67
Tt.Ar (mm <sup>2</sup> )	694.1 (110.3)	727.3 (80.4)	0.32	703.1 (133.1)	700.8 (93.5)	0.92
Ct.Ar (mm <sup>2</sup> )	127.1 (23.8)	111.1 (12.1)	0.02	120.9 (24.1)	110.2 (20.6)	0.01
Tb.Ar (mm <sup>2</sup> )	567.0 (114.3)	616.2 (81.4)	0.16	582.2 (140.2)	590.6 (90.8)	0.72
Ultimate stress (MPa)	27.8 (8.3)	19.2 (4.9)	<0.001	25.9 (8.6)	21.2 (7.0)	0.003
Load cortex distal (%)	37.5 (9.5)	36.0 (7.5)	0.59	36.1 (9.8)	35.5 (7.8)	0.72
Load cortex prox (%)	70.1 (9.5)	72.8 (9.3)	0.40	69.2 (12.1)	70.9 (9.6)	0.40

nonlinear boundaries between the cases [29], and notably, they are stable and require minimal tuning of parameters [30]. This allows SVMs to work well in many applications and reduces overfitting to the specific dataset. Other machine learning methods such as GBM may provide higher accuracy of classification but are computationally expensive

and, most importantly, require extensive tuning of parameters. SVMs have an important advantage in that they find a global, unique solution unlike other methods such as artificial neural networks that may find local minima [31]. This advantage of SVMs potentially makes this approach more robust for application to diverse clinical datasets.

**Table 3** DXA parameters [mean (SD)] for the fracture and control groups

DXA	Control (n=28)	Forearm low-trauma fx (n=14)	p value	Control (n=88)	Any low-trauma fx (n=44)	p value
FN aBMD (g/cm <sup>2</sup> )	0.711 (0.101)	0.667 (0.069)	0.15	0.700 (0.120)	0.654 (0.091)	0.03
LS aBMD (g/cm <sup>2</sup> )	0.961 (0.133)	0.833 (0.123)	0.01	0.935 (0.152)	0.861 (0.114)	0.01
UDR aBMD (g/cm <sup>2</sup> )	0.363 (0.049)	0.319 (0.065)	0.02	0.348 (0.068)	0.334 (0.055)	0.24

**Table 4** SVM classification results for the forearm fractures and for fractures at all sites. Five different models were used for each fracture group. The first was HR-pQCT, FE, and DXA parameters; the second was HR-pQCT and FE parameters; the third was FE parameters alone; the fourth was DXA parameters including simulated ultradistal radius

	Forearm fractures					All fractures				
	HR-pQCT FE DXA	HR-pQCT FE	FE	LS, FN, UDR DXA	LS, FN DXA	HR-pQCT FE DXA	HR-pQCT FE	FE	LS, FN, UDR DXA	LS, FN DXA
Accuracy (%)	81.0	83.3	78.6	71.4	59.5	68.9	68.9	69.7	56.8	50.0
Sensitivity (%)	68.8	73.3	63.2	55.6	43.5	52.8	52.9	53.7	42.0	37.5
Specificity (%)	88.5	88.9	91.3	83.3	78.9	79.7	79.0	80.8	80.4	75.0
ROC-AUC	0.80	0.82	0.80	0.71	0.63	0.68	0.67	0.69	0.62	0.56

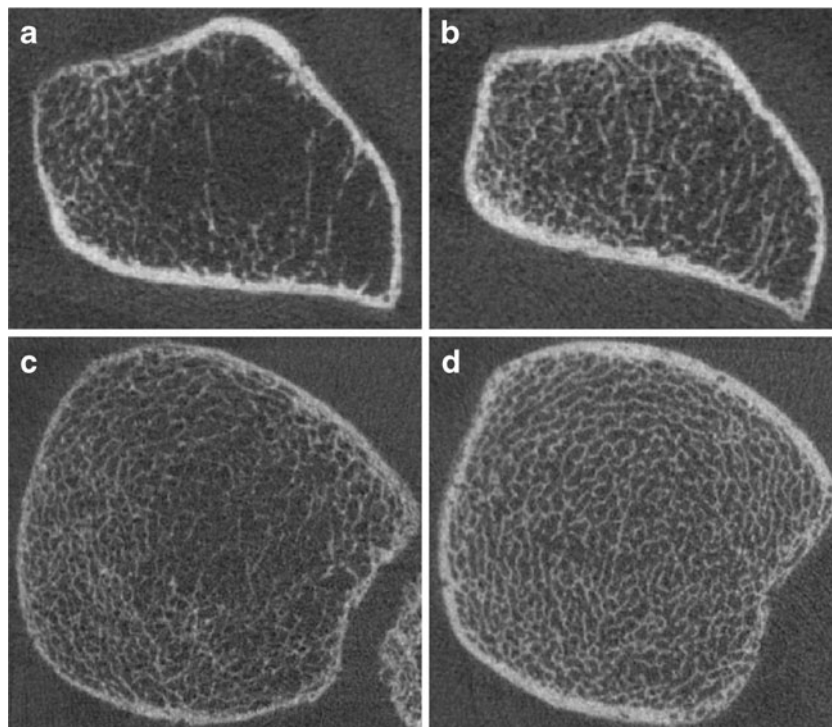
(UDR); and the last was lumbar spine (LS) and femoral neck (FN) DXA only. Accuracy, sensitivity, specificity, and the receiver operating characteristic area under the curve (ROC-AUC) are shown for each model

To date, only one other study has used machine learning techniques to classify postmenopausal women with and without a history of low trauma [12]. Similar to our findings, Atkinson et al. reported more accurate classification when bone density, geometry, and microstructural outcomes obtained with additional imaging modalities (QCT and HR-pQCT) were included in their GBM model in addition to standard DXA-derived aBMD, indicating the importance of microarchitectural parameters. In addition, whereas not all bone outcomes were significantly different between groups when compared individually, small contributions of each variable combined into a single model allowed for better fracture discrimination. Atkinson et al. reported higher AUC's for predicting forearm fracture cohorts compared with our current study, which was possibly due to

their larger sample size ( $n=99$  forearm fractures) [12]. While our study is smaller, we used a tenfold cross-validation scheme, which gives an excellent indication of how our model would perform in practice on a new dataset.

Previous studies used PCA and logistic regression models to determine if HR-pQCT measures of bone microarchitecture and FE estimates of bone strength could discriminate between postmenopausal women with and without previous fracture [9–11]. These analyses differ from our approach in that they attempt to determine the individual parameters (or principal components of parameters) that best discriminate fracture cases from controls, whereas the goal of SVM is to perform classification. Despite the different analytic approaches and goals, results of these studies consistently indicate that HR-pQCT measurements of bone microarchitecture and FE

**Fig. 1** Representative scans of a participant with a low trauma forearm fracture and an age-matched nonfracture control at the distal radius (a and b, respectively) and at the distal tibia (c and d, respectively)



analysis estimates of bone strength outperform DXA-measured aBMD when discriminating between postmenopausal women with and without a history of low trauma fracture. Consistent with our findings, Vilayphiou et al. showed that FE-estimated bone stiffness had high contributions to the principal components [10]. The relatively similar performance between our model using HR-pQCT, DXA, and FE parameters and the FE parameters alone indicate that the FE measurements provides an excellent “summary” measurement of bone quality.

Clinical DXA scans of the forearm are rare; therefore, we included models based on DXA measurements of LS and FN with and without UDR aBMD. The model with only LS and FN aBMD represents a typical clinical scenario and, as expected, resulted in the weakest classification as shown in Table 4. Based on the model that included UDR, FN, and LS aBMD, it is important to include a scan of the distal radius to classify forearm fractures. Our data may actually overestimate the importance of a UDR scan because our simulated UDR aBMD measurements were focused on precisely the area that is most likely to fracture, whereas a real UDR DXA would cover a broader region and may be less sensitive to classifying forearm fracture. It is not surprising that models including UDR aBMD more accurately classify forearm fractures. However, HR-pQCT measurements still outperformed all DXA measurements when classifying forearm fractures.

We found that the SVMs were better at discriminating low-trauma forearm fractures compared with low-trauma fractures at other sites. This is likely attributable to site specificity since the HR-pQCT scans obtained at the distal radius were used to classify the forearm fractures. The difference in discriminative ability may also be due to the large variation in fracture types in our study. We had few participants ( $n=6$ ) with fractures at “major” osteoporotic sites (i.e., hip and spine). It is possible that these fractures would be better predicted by the SVMs because women who suffer these types of fractures tend to be at more advanced disease stages compared with women who sustain forearm fractures [32]. A larger cohort of fracture cases would strengthen our understanding of the potential for SVMs to classify fractures. However, our current study indicates that there is excellent potential for this method in classification studies of all low-trauma fractures.

A limitation of our study is that it is a cross-sectional design and has a relatively small sample size, and the acquisition of HR-pQCT scans were done retrospectively after the fractures were sustained. Despite these weaknesses, our study provides strong support for SVM combined with HR-pQCT to classify fractures better than by the clinical gold standard of DXA aBMD, even at the UDR. It would be ideal to expand this research with a larger fracture cohort and to perform the study prospectively, and work in this area

is currently underway at our laboratory and at other sites worldwide. There are also some technical limitations with our study, the first being that the FE estimates are based on a single uniaxial compression test. While this is currently the gold standard for FE applied to HR-pQCT, it may be advantageous to expand the number of simulated FE tests for a more comprehensive assessment of bone strength, and hence better prediction of the strength of the bone during a fall on an outstretched arm. In addition, FE models were computed using a homogenous Young's modulus for each element. Strength estimates using this method are highly correlated with experimental estimates of bone strength [7]; however, FE measurements may be affected by mineralization differences. In future, it may be important to account for such differences using scaled modulus values based on density measurements.

The application of SVMs to HR-pQCT data shows great promise to classify postmenopausal women with and without a history of low-trauma fracture. Our results suggest that this is an accurate classification method for forearm fractures and has potential to be applicable for classification of fractures at other sites. This is the first study to apply SVM to HR-pQCT data and is strengthened by incorporating cross-validation for testing and including FE estimates of bone strength.

**Acknowledgments** We wish to acknowledge the ongoing efforts of the national Canadian Multicentre Osteoporosis Study (CaMos). We would also like to thank Ms. Irene Hanley and Ms. Shannon Boucousis for their assistance with scan acquisition and Ms. Jane Allan and Ms. Bernice Love for their assistance with participant recruitment. We are grateful to all of the participants who volunteered for the study.

**Sources of funding** This study was supported by Canadian Institutes of Health Research and Vanier Canada Graduate Scholarships.

**Conflicts of interest** None.

## References

1. Black DM, Cummings SR, Genant HK, Nevitt MC, Palermo L, Browner W (1992) Axial and appendicular bone density predict fractures in older women. *J Bone Miner Res* 7:633–638
2. Schuit CE, van der Klift M, Weel A, de Laet C, Burger H, Seeman E, Hofman A, Uitterlinden G, van Leeuwen J, Pols AP (2004) Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone* 34:195–202
3. van der Linden JC, Weinans H (2007) Effects of microarchitecture on bone strength. *Curr Osteoporos Rep* 5:56–61
4. Sornay-Rendu E, Boutroy S, Munoz F, Delmas PD (2007) Alterations of cortical and trabecular architecture are associated with fractures in postmenopausal women, partially independent of decreased BMD measured by DXA: the OFELY study. *J Bone Miner Res* 22:425–433
5. Boutroy S, Bouxsein ML, Munoz F, Delmas PD (2005) In vivo assessment of trabecular bone microarchitecture by high-resolution

- peripheral quantitative computed tomography. *J Clin Endocrinol Metab* 90:6508–6515
6. MacNeil JA, Boyd SK (2007) Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med Eng Phys* 29:1096–1105
  7. MacNeil JA, Boyd SK (2008) Bone strength at the distal radius can be estimated from high-resolution peripheral quantitative computed tomography and the finite element method. *Bone* 42:1203–1213
  8. Melton LJ, Riggs BL, van Lenthe GH, Achenbach SJ, Müller R, Bouxsein ML, Amin S, Atkinson EJ, Khosla S (2007) Contribution of in vivo structural measurements and load/strength ratios to the determination of forearm fracture risk in postmenopausal women. *J Bone Miner Res* 22:1442–1448
  9. Boutroy S, Van Rietbergen B, Sornay-Rendu E, Munoz F, Bouxsein ML, Delmas PD (2008) Finite element analysis based on in vivo HR-pQCT images of the distal radius is associated with wrist fracture in postmenopausal women. *J Bone Miner Res* 23:392–399
  10. Vilaythiou N, Boutroy S, Sornay-Rendu E, Van Rietbergen B, Munoz F, Delmas PD, Chapurlat R (2010) Finite element analysis performed on radius and tibia HR-pQCT images and fragility fractures at all sites in postmenopausal women. *Bone* 46:1030–1037
  11. Vico L, Zouch M, Amirouche A, Frère D, Laroche N, Koller B, Laib A, Thomas T, Alexandre C (2008) High-resolution pQCT analysis at the distal radius and tibia discriminates patients with recent wrist and femoral neck fractures. *J Bone Miner Res* 23:1741–1750
  12. Atkinson EJ, Thorneau TM, Melton LJ, Camp JJ, Achenbach SJ, Amin S and Khosla S (2012) Assessing fracture risk using gradient boosting machine (GBM) models. *J Bone Miner Res*
  13. Macdonald HM, Nishiyama KK, Kang J, Hanley DA, Boyd SK (2011) Age-related patterns of trabecular and cortical bone loss differ between sexes and skeletal sites: a population-based HR-pQCT study. *J Bone Miner Res* 26:50–62
  14. Nishiyama KK, Macdonald HM, Buie HR, Hanley DA, Boyd SK (2010) Postmenopausal women with osteopenia have higher cortical porosity and thinner cortices at the distal radius and tibia than women with normal aBMD: an in vivo HR-pQCT study. *J Bone Miner Res* 25:882–890
  15. Kreiger N, Tenenhouse A, Joseph L, MacKenzie T, Poliquin S, Brown J, Prior J, Rittmaster R (1999) The Canadian Multicentre Osteoporosis Study (CaMos): background, rationale, methods. *Can J Aging* 18:376–387
  16. Berger C, Langsetmo L, Joseph L, Hanley DA, Davison KS, Josse RG, Prior JC, Kreiger N, Tenenhouse A, Goltzman D (2009) Association between change in BMD and fragility fracture in women and men. *J Bone Miner Res* 24:361–370
  17. Laib A, Häuselmann HJ, Rüegeegger P (1998) In vivo high resolution 3D-QCT of the human forearm. *Technol Health Care* 6:329–337
  18. Sekhon K, Kazakia GJ, Burghardt AJ, Hermannsson B, Majumdar S (2009) Accuracy of volumetric bone mineral density measurement in high-resolution peripheral quantitative computed tomography. *Bone* 45:473–479
  19. MacNeil JA, Boyd SK (2008) Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med Eng Phys* 30:792–799
  20. Buie HR, Campbell GM, Klinck RJ, MacNeil JA, Boyd SK (2007) Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone* 41:505–515
  21. Burghardt AJ, Kazakia GJ, Link TM, Majumdar S (2009) Automated simulation of areal bone mineral density assessment in the distal radius from high-resolution peripheral quantitative computed tomography. *Osteoporos Int* 20:2017–2024
  22. Müller R, Rüegeegger P (1995) Three-dimensional finite element modelling of non-invasively assessed trabecular bone structures. *Med Eng Phys* 17:126–133
  23. Van Rietbergen B, Weinans H, Huiskes R, Odgaard A (1995) A new method to determine trabecular bone elastic properties and loading using micromechanical finite-element models. *J Biomech* 28:69–81
  24. Chiu J, Robinovitch SN (1998) Prediction of upper extremity impact forces during falls on the outstretched hand. *J Biomech* 31:1169–1176
  25. Hayes WC, Piazza SJ, Zysset PK (1991) Biomechanics of fracture risk prediction of the hip and spine by quantitative computed tomography. *Radiol Clin North Am* 29:1–18
  26. Keaveny TM, Bouxsein ML (2008) Theoretical implications of the biomechanical fracture threshold. *J Bone Miner Res* 23:1541–1547
  27. Vapnik V Estimation of dependences based on empirical data: Springer series in statistics. Springer, New York
  28. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *SIGKDD Explorations* 11
  29. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
  30. Valyon J, Horváth G (2003) A weighted generalized LS-SVM. *Period Polytech Electr Eng* 47:229–251
  31. Suykens JAK, De Brabanter J, Lukas L, Vandewalle J (2002) Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* 48:85–105
  32. Harvey N, Dennison E, Cooper C, ASBMR (2009) Chapter 38. Epidemiology of osteoporotic fractures. In: Rosen JC (ed) *Primer on the metabolic bone diseases and disorders of mineral metabolism*, 7th edn. American Society for Bone and Mineral Research, Washington, pp 197–203