ORIGINAL ARTICLE

Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures

A. Valentinitsch[1] · S. Trebeschi[1] · J. Kaesmacher[1] · C. Lorenz[2] · M. T. Löffler[1] · C. Zimmer[1] · T. Baum[1] · J. S. Kirschke[1]

Abstract

Summary Our study proposed an automatic pipeline for opportunistic osteoporosis screening using 3D texture features and regional vBMD using multi-detector CT images. A combination of different local and global texture features outperformed the global vBMD and showed high discriminative power to identify patients with vertebral fractures.

Introduction Many patients at risk for osteoporosis undergo computed tomography (CT) scans, usable for opportunistic (non-dedicated) screening. We compared the performance of global volumetric bone mineral density (vBMD) with a random forest classifier based on regional vBMD and 3D texture features to separate patients with and without osteoporotic fractures.

Methods In total, 154 patients (mean age 64 ± 8.5, male; $n = 103$) were included in this retrospective single-center analysis, who underwent contrast-enhanced CT for other reasons than osteoporosis screening. Patients were dichotomized regarding prevalent vertebral osteoporotic fractures (noFX, $n = 101$; FX, $n = 53$). Vertebral bodies were automatically segmented, and trabecular vBMD was calculated with a dedicated phantom. For 3D texture analysis, we extracted gray-level co-occurrence matrix Haralick features (HAR), histogram of gradients (HoG), local binary patterns (LBP), and wavelets (WL). Fractured vertebrae were excluded for texture-feature and vBMD data extraction. The performance to identify patients with prevalent osteoporotic vertebral fractures was evaluated in a fourfold cross-validation.

Results The random forest classifier showed a high discriminatory power (AUC = 0.88). Parameters of all vertebral levels significantly contributed to this classification. Importantly, the AUC of the proposed algorithm was significantly higher than that of volumetric global BMD alone (AUC = 0.64).

Highlights

- We present an automatic pipeline for opportunistic osteoporosis screening using 3D texture features and regional vBMD.
- A combination of local texture and density features performed better than global vBMD alone to distinguish between patients with and without osteoporotic vertebral fractures.
- All vertebral levels were important to distinguish between patients with and without osteoporotic vertebral fractures.

✉ A. Valentinitsch
alexander.valentinitsch@tum.de

S. Trebeschi
stefano.trebeschi@tum.de

J. Kaesmacher
johannes.kaesmacher@tum.de

C. Lorenz
cristian.lorenz@philips.com

M. T. Löffler
m.loeffler@tum.de

C. Zimmer
claus.zimmer@tum.de

T. Baum
thomas.baum@tum.de

J. S. Kirschke
jan.kirschke@tum.de

[1] Department of Diagnostic and Interventional Neuroradiology, Klinikum rechts der Isar, Technische Universität München, München, Germany

[2] Philips Research Hamburg, Hamburg, Germany

Conclusion The presented classifier combining 3D texture features and regional vBMD including the complete thoracolumbar spine showed high discriminatory power to identify patients with vertebral fractures and had a better diagnostic performance than vBMD alone.

Abbreviations

DXA	dual-energy X-ray absorptiometry
MDCT	Multi-detector computed tomography
vBMD	Volumetric bone mineral density
HAR	Haralick features
HOG	Histogram of gradients
LBP	Local binary pattern
WL	Wavelets
RF	Random forest

Introduction

Osteoporosis is a potentially devastating disease associated with bone mineral loss and deterioration of the delicate bony microstructure especially vertebral and hip fractures which are associated with high mortality and morbidity [1]. As many patients are often not diagnosed with osteoporosis prior to osteoporotic fractures, routinely identifying patients at risk is desirable [2]. Vertebral fractures are the second most common osteoporotic fractures [3]. They are associated with low bone mineral density (BMD), which is routinely assessed by dual-energy X-ray absorptiometry (DXA) [4]. However, DXA measures only areal BMD, which cannot distinguish between degenerative changes, cortical and trabecular bone, cannot assess the three-dimensional (3D) shape of each vertebra, and may overestimate BMD in obese subjects. Thus, poor accuracy to predict osteoporotic fractures has been reported [5]. This implicates an urgent need to develop a clinically feasible tool that can improve fracture risk assessment at the spine. Quantitative computed tomography (qCT) is an established alternative allowing for 3D assessment of bone mineral density [6]. Based on such data, finite element analysis and biomechanical features have already been used to improve the performance in fracture risk assessment, differentiating individuals with and without prevalent vertebral fractures [7], and predicting incidental vertebral fractures [5]. Despite the availability of 3D data, only two-dimensional (2D) texture analysis techniques have been applied to CT images in an in vivo setting [8]. Data mining techniques such as feature extraction (i.e., texture, shape, density, stiffness, etc.) that utilizes the full available 3D information of the vertebral composition, is expected to further enhance the diagnostic accuracy by combining machine learning and statistical analysis intelligently. The combination of these techniques may further enhance the discrimination of patients with and without fractures. Compared to DXA, qCT is affiliated with a substantially higher radiation dose, what limits the broad use as a screening technique up to now [9]. On the other hand, there are many abdominal CT scans available of patients at risk obtained for other indications, which can be used for "opportunistic screening", without additional exposure and substantial costs. Recently, such computed tomography (CT) scans, partly or completely covering the spine, were used to identify patients with osteoporosis, detect, and predict vertebral compression fractures from reconstructed sagittal images [6, 10, 11].

In this feasibility study, we evaluated an advanced automatic algorithm for opportunistic osteoporosis screening in nondedicated CT images. In detail, we developed a quantitative method for the identification of patients with prevalent osteoporotic vertebral fractures in existing CT images using a random forest classifier that uses 3D texture features in combination with a global and local volumetric BMD.

Materials and methods

Human subjects and MDCT imaging

Ethics approval was obtained from the local ethics committee (11/5022A1). Due to the retrospective nature of the study, the need for informed consent was waived. Retrieved from our local database, we reviewed consecutive patients who received MDCT, in the time between February 2007 and February 2008, for reasons of cancer staging, restaging, or follow-up after surgical treatment or chemotherapy.

Inclusion criteria for the present study consisted of (1) patients older than 38, (2) a CT scan of the thoracolumbar spine including sagittal reformations, (3) a bone mineral phantom within the scan field, and (4) the absence of any diseases affecting the spine such as bone metastases, hematological disorders, or metabolic bone diseases other than osteoporosis. To definitively exclude spinal metastasis, we included only patients with available follow-up scans of the spine confirming the absence of bone metastases. In total, 154 patients were included in the study (males; $n = 103$ and females; $n = 51$). These oncologic patients had histologically proven neoplasms of the gastrointestinal tract (102), lymphatic system (20), urinary tract (8), respiratory tract (6), sarcoma (7), or

other solid tumors (11). The majority of patients showed no signs of distant metastasis (92); a minority were lymphoma patients (20); in the remaining cases non-spinal, distant metastases were present (42). Due to the fact that all subjects underwent screening for cancer metastasis, intravenous contrast medium (Imeron 400; Bracco, Konstanz, Germany) was administered using a high-pressure injector (Fresenius Pilot C; Fresenius Kabi, Bad Homburg, Germany). Intravenous contrast medium injection was performed with a delay of 70 s, a flow rate of 3 ml/s, and a body weight–dependent dose (80 ml for body weight up to 80 kg, 90 ml for body weight up to 100 kg, and 100 ml for body weight over 100 kg). Furthermore, all patients received 1000 ml oral contrast medium (Barilux Scan; Sanochemia Diagnostics, Neuss, Germany). All images were acquired with a Siemens CT scanner (Somatom 128, Siemens Healthcare AG, Erlangen, Germany) with calibration phantom with two rods (Osteo Phantom, Siemens Healthcare AG, Erlangen, Germany).

A patient was diagnosed with established osteoporosis (FX) if an osteoporotic vertebral fracture was detected in the image (53 patients). According to the semiquantitative Genant classification, vertebrae with a height loss of more than 20% (grade 1) and the typical morphology of osteoporotic fractures were considered as fractured [12]. A total of 101 patients had no signs of osteoporotic vertebral fractures (noFX).

Bone mineral density

The calibration phantom values were used for Hounsfield units (HU) to vBMD conversion. To account for the contrast medium administered to all subjects, a linear conversion factor for portal-venous (PV) was applied ($BMD_{QCT} = 1.02 \times BMD_{MDCT} - 18.72$ mg/ml), as proposed in [13]. The corrected vBMD value for each vertebra of each patient was computed by sampling all voxels within the respective trabecular compartment. Finally, the vBMD value of the thoracic, lumbar, and thoracolumbar spine was determined by averaging the mean vBMD values and standard deviation (SD) of their respective vertebrae. Additional to the global mean for each vertebral level, we extracted also skewness and kurtosis, which we refer to as global density features (BMD) for classification.

Global and local feature extraction

We extracted features on a global (i.e., vBMD) and local level (i.e., regional). Global features were extracted for the complete eroded vertebral body. Both density calculation and texture analysis were performed using the calibrated scans. Due to the linear conversion used for calibration, internal microarchitectures and morphological patterns described by the textural features remained independent from this calibration. To fully utilize the advantage of texture analysis locally, we

defined 27 subregions as proposed by [14] of each vertebra of our spine template (TLSSM16) generated in [15]. The center of the largest sphere fitting in the vertebral mask was defined as the center point of the vertebral body. Additionally, we extracted surface points of the vertebral endplates (i.e., superior and inferior endplate points), which we projected to the center point. The given set of 3D points was used to compute the plane that best fits those points by minimizing the sum of the quadratic distance (perpendicular to the plane) between the plane and the points. The fit was performed by computing the eigenvectors associated with the distribution of the points. Using a combination of two eigenvectors as the orthonormal basis of the planes, we extracted three distinct planes: superior-inferior plane (i.e., fitted plane), anterior-posterior plane, and medial-lateral plane. We divided the largest fitted sphere into three parts to define superior (S), mid-transverse (T), and inferior (I) regions using the fitted transverse plane, and into lateral (L) and medial (M) regions using the defined sagittal plane. Coronally, the vertebral bodies were divided into thirds to define the anterior (A), mid-coronal (C), and posterior (P) region using the anterior-posterior plane. The posterior elements were separated from the vertebral body using the anterior-posterior plane fitted to the posterior border of the vertebral body, i.e., the anterior border of the spinal canal. This intersection resulted in 27 subregions, which are depicted in Fig. 1. We extracted density (regional volumetric bone mineral density (BMDr)) and texture features for each vertebra for all defined subregions using different texture analysis techniques. We computed simple statistical descriptors for those features using the mean, standard deviation, skewness, and kurtosis.

Pre-processing

Each vertebra of the thoracolumbar spine was localized and segmented by an automated algorithm based on shape model matching [16]. The corresponding vertebra of the spine template (TLSSM16) was then aligned to the segmented vertebra to define the vertebral subregions for texture analysis. More specifically, we first estimated a rigid motion (i.e., rotation and translation) which roughly aligned the TLSSM16 to the sample vertebra. Next, we fitted the vertebral body of the TLSSM16 to the vertebral body of the sample via affine transform, which adds anisotropic scaling. Once the registration pipeline was concluded, we could easily warp the defined subregions to the sample vertebra. To exclude the surrounding cortical shell and limit the analysis to the trabecular compartment, we eroded the resulting mask of the vertebral body by a sphere with a radius of 4 voxels.

The implementation of the registration procedure was based on the elastix framework [17]. Visual inspection has been conducted on the results of both segmentation and registration to check the accuracy of the intermediate results. In
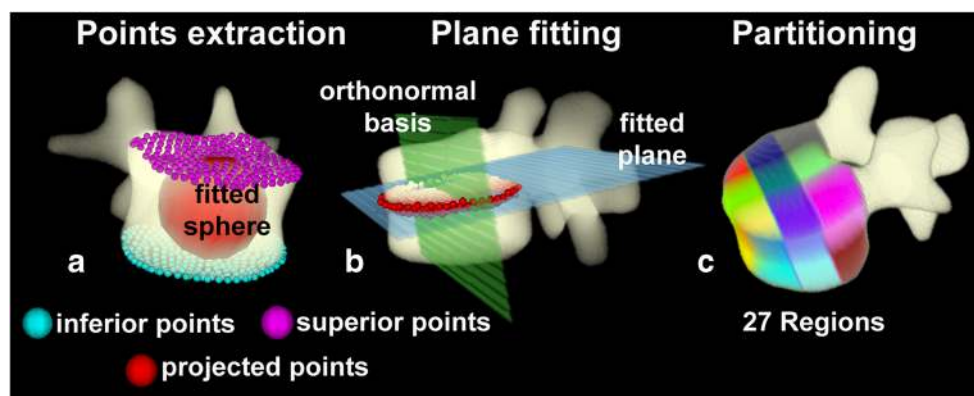
Fig. 1 Region definition process. (**a**) The biggest sphere, fitting in the mask defined the center point of the vertebral body. Additionally, we extracted surface points of the vertebral endplates, which we projected to the center point. (**b**) The given set of 3D points was used to compute the three orthogonal planes: superior-inferior plane (i.e., fitted plane), anterior-posterior plane, and medial-lateral plane. (**c**) The intersections resulted in 27 regions

total, 11 vertebrae had to be excluded from the procedure due to incorrect segmentation ($n = 9$) or registration ($n = 2$). The reason for this failure seemed to be high-grade fractures ($n = 6$) or severe degeneration in fractured vertebrae ($n = 3$) and abnormalities of the posterior elements ($n = 2$).

Three-dimensional textures analysis

Haralick features of the 3D co-occurrence matrix (HAR)

The Haralick features (HAR) are a set of features computed on the gray-level co-occurrence matrix (GLCM), a joint histogram of which the elements describe the occurrence of two intensity levels of being neighbors at a certain offset [18]. The algorithm for the gray-level co-occurrence matrix used in this work was set to the following parameters: 16 bins, offset of 1, in 13 distinct directions which defined the GLCM. Thirteen different HAR were used, which are reported in the supplemental material and described in [3, 19]. However, the vicinity of 2 voxels is not uniquely defined. An element lying in the 3D space has six direct neighbors with whom it shares one face and 20 semi-direct neighbors, which result in 13 unique directions. To address such directional ambiguity, we compute the mean and standard deviation of the Haralick features (HAR) in each possible direction. These are called the *angular mean* and *angular standard deviation*, respectively [19]. Both the angular mean and standard deviation vectors were computed as descriptors of the textures in a region.

3D histograms of oriented gradients

Histograms of oriented gradients (HOG) [20] describe textural patterns based on the gradient information. The gradient of a volume is defined at a voxel v as the change of intensity between the neighbors of v in the axial, sagittal, and coronal planes. The difference in intensity in each direction generates

a vector called *gradient vector*. Such a vector is computed for each voxel v. To compute HOG features, the gradient vector is projected on the 20 faces of an icosahedron (i.e., a 20-sided dice) built around the voxel v [20]. Each normalized projection generates a vector, the magnitude of which is binned in a histogram. The textural descriptor was estimated by summing over the histograms in a certain region. Additionally, the same procedure can be applied to the gradient itself, obtaining in this way the descriptors of second-order gradients.

3D local binary patterns

Local binary patterns (LBP) were first introduced in 2D [21] as a way to uniquely identify the specific displacement of intensities around a pixel, with the main advantage of being invariant to rotations. The original procedure comprised the readout of the intensity values around a circle centered on the pixel of interest in a binary fashion. If the surrounding pixel, value is bigger than the central pixel, it gets the value of 1 and otherwise 0. The extension to a 3D space required the development of a more complex procedure to readout values from a sphere surrounding a certain voxel and describe them in a compact and unique fashion. Such a procedure is based on *spherical harmonics*, a mathematical framework, which allows the approximation of functions defined on a sphere [22]. Additionally, to confer to the descriptor's rotational invariance, as originally proposed in 2D, the kurtosis was computed on the distribution of sampled voxels. This resulted in feature maps for each voxel location to a higher dimensional vector representing the particular 3D texture surrounding the voxel. Two parameters were set for this descriptor: the radius or the sampling sphere $r = 2$, 3, and 4 voxels and the number of coefficients $f = 3$ used by the spherical harmonics. The higher the number of coefficients, the more patterns and textures can be represented.

The most direct way to use LBP for the analysis of textures in a region would be to look for the most common pattern in that region. However, this approach is sensitive to noise, which changes the coefficients of the higher frequencies. By clustering these vectors according to their similarity, we were less sensitive to noise [23]. More specifically, we clustered the extracted 3D LBP features using k-means with $k = 2, 3$, and 4. Each resulting cluster, represented by its respective mean, was used as a descriptor, along with its cardinality.

3D wavelet decomposition

The term *wavelet* refers to a signal having a wave-like oscillation with amplitude that increases from zero up to a certain value and then decreases back to zero. Similar to sinusoidal functions in classical Fourier analysis, wavelets can be used as a basis function in the decomposition of a complex signal [24]. Unlike Fourier analysis, however, the limited support of wavelets easily allows the modeling of local frequency variations (or textures, in the case of images).

More specifically, a discrete 3D signal (i.e., the CT image) is decomposed into the weighted sum of a high-frequency signal (H) and a lower one (L) in each direction. This procedure generates eight sub-bands of one-eighth the size of the original volume (HHH, HHL, HLH, HLL, LHH, LHL, LLH, and LLL), one for each combination of the type of frequency and dimension applied. High frequency coefficients capture high-frequency signals such as edges and noise, whereas low frequency coefficients give a smoother representation of the signal. The combination of the high and low frequency highlights edges and ridges in specific directions as indicators of textures.

In addition, wavelet decomposition implicitly offers a multiresolution approach by recursively applying the decomposition on the LLL sub-band.

We used simple statistical descriptors (i.e., mean, standard deviation, skewness, and kurtosis) on each sub-band for two subsequent resolution levels [25].

Classification

Among all classification algorithms presented in the literature, we opted for random forests (RFs) [26]. Random forests are an ensemble of different decision trees built on random subsets of the input space. A decision tree is a multivariate classifier, which splits multidimensional data recursively, one variable at the time, to create homogeneous subsets of data. The classification of new samples is performed by assigning the class of the subset the new samples falls into. Assembling multiple decision trees together creates a random forest, which offers higher robustness to noise and higher generalization compared to a single decision tree. We used 2001 trees. To avoid overfitting, our RFs implement decision trees were built on a random subset of the input space [27]. Such RFs have been shown to be efficient classifiers, able to handle complex and non-linear classification problems and large and high-dimensional datasets and provide high accuracy [28]. Its training is performed using a local optimal strategy which recursively minimizes the probability of a random sample to be misclassified, a.k.a. Gini index. A reduction of the Gini index given by the selection of a certain feature, summed over all decision trees in the forest, a.k.a. Gini importance (GI), provides a quantification of the importance of each feature during the classification task [26].

At this point, we built the input space (i.e., feature vector) used for the classification. Specifically, we extracted textural features according to the section Three-dimensional Textures Analysis from each vertebral body (global) and the BMD mean, standard deviation, skewness, and kurtosis from a global and local level (i.e., the 27 regions). Subsequently, feature vectors were concatenated for each vertebra in the thoracolumbar spine. Seventy-nine vertebrae with existing fractures (as well as 2 vertebrae with incorrect segmentations without fracture) were excluded from the analysis to avoid bias. These missing values were replaced by the sample mean.

Finally, since textures could be hampered by noise, but also may be destroyed by smoothing, we computed each feature on four increasing levels of Gaussian smoothing. Specifically, we applied a Gaussian isotropic kernel sigma = 0, 1/3, 2/3, and 1—where 0 is no smoothing—and sized three times the sigma.

Feature selection

Reducing the input space to the most relevant features, a.k.a. feature selection, can improve the results significantly, especially in this case, where the information contained in one vertebra could likely be correlated to adjacent vertebrae causing information redundancy. To identify the most important features, we opted for an exponential search: from the training procedure, we extracted the GI and ranked the features accordingly. Then we re-ran the training using the first m features, where $m = 2, 4, 8, \ldots 32{,}768$ (in a 2^n fashion). A quadratic function was used to model the change in performance w.r.t. n. The vertex of the parabola was used as optimal cut.

Statistical analysis

A significant level of 0.05 was used in all statistical analysis. Descriptive statistics were given by means and standard deviations (SD), after checking for normal distribution. To compare the global density in patients with fractures (FX) and patients without fractures (noFX), we used a Student's t test. We used a pairwise Pearson correlation coefficient (r) to investigate the relationship of vBMD against age.

The fracture classification performance was computed on a fourfold cross-validation, repeated 10 times with a random forest of 2001 trees, classifying if the patient was in the FX or noFX group (i.e., binary classification). More specifically, the original dataset (i.e., sample) is randomly partitioned into four equal size subsamples. Of the four subsamples, a single subsample is retained as the validation data for testing the model, and the remaining three subsamples are used as training data. This fourfold cross-validation was repeated 10 times with different randomly chosen subsamples to account for possible differences between subsequent trainings. To assess the diagnostic capability of single features as well as the whole model, receiver operator characteristic (ROC) curve analysis was used. The AUC comparisons were statistically tested using the McNiel method.

Results

Patient statistics

Patient characteristics are depicted in Table 1. There was no statistically significant difference in age between the FX group (66.6 ± 9.2 years) and the noFX group (63.5 ± 7.9 years; $p > 0.08$). No significant difference was observed between females and males in terms of age neither as a whole nor in each subgroup. Overall, 79 fractures (Genant grade 1; $n = 20$, grade 2; $n = 40$, grade 3; $n = 19$) were observed in 53 patients. Seventeen patients had multiple fractures. The most common fracture location was the transition between the thoracic and lumbar spine (T11-L2), observed in 36 patients. The middle thoracic spine (T6-T8) was also a common fracture location site (31 patients).

In the studied cohort, the global bone mineral density of FX patients (86.5 ± 19.8 mg/cm^3) was significantly lower ($p < .001$) than the mineral density of the noFX patients (103.8 ± 23.8 mg/cm^3). Furthermore, within the FX group, female subjects presented with a trend towards a lower BMD in the thoracic and lumbar spine as compared to males ($p = 0.07$). Such a difference was also observed in the noFX group. Only weak negative correlations were detected between vBMD and age ($r = -0.26$, $p < 0.01$). The distribution of mean BMD of all vertebrae in the thoracic and lumbar spine is displayed in Fig. 2.

Classification

The overall classification performance was 0.88 AUC on a fourfold cross-validation via feature selection where the performance function reached its global maximum at $2^{7.8} \approx 2^8 = 256$ features (Fig. 3a). The performance decreased to AUC of 0.71 when using the entire input space. AUC comparison analysis showed that a combination of important features significantly ($p < 0.01$) outperformed individual features.

Figure 3b shows the mean GI of each feature class as computed by the RF. LBP and regional BMD (BMDr) are highlighted as the most relevant parameters, accounting for the highest cumulative GI (i.e., $> 50\%$). Global vBMD showed the least importance. On the other hand, regional parameters were important in all vertebral levels, i.e., there was no region with unimportant information (Fig. 3c). Regional density (BMDr) was the most discriminative factor within T3–5, whereas the whole thoracolumbar spine was dominated by structural features (texture) such as LBP and WL. The low importance of global features like vBMD was reflected in an AUC of only 0.64 (Table 2). The comparison of the receiver operating characteristic (ROC) is depicted in Fig. 3d. Figure 4 shows the computation of the most important feature (LBP), representing the L1 of a FX patient compared to the L1 of a noFX subject.

Finally, we did not observe a significant improvement in performance (AUC) for any smoothing setting applied (data not shown).

Table 1 Patient age (in years) and volumetric bone mineral density (vBMD, in mg/cm^3) of the lumbar and thoracic spine, presented as minimum (min), maximum (max) and mean ± standard deviation (SD)

	Age					vBMD (thoracic)		vBMD (lumbar)	
	n	min	max	mean	SD	mean	SD	mean	SD
FX	53	44	82	66.6	9.2	88.22	20.89	81.51	20.69
FX (M)	35	44	78	68.8	8.3	91.97	22.39	83.63	19.93
FX (F)	18	47	82	65.5	10.5	80.91	15.14	77.40	21.52
noFX	101	39	88	63.5	7.9	105.59	25.74	98.95	23.09
noFX (M)	68	42	88	62.2	7.1	104.25	28.11	98.32	24.12
noFX (F)	33	39	74	64.2	9.2	108.36	19.69	100.24	20.73
All	154	39	88	64.6	8.5	99.61	25.55	92.95	23.78

Values are given for all patients, the fracture (FX) and non-fracture subgroups (noFX), divided by gender (M: male; F: female)
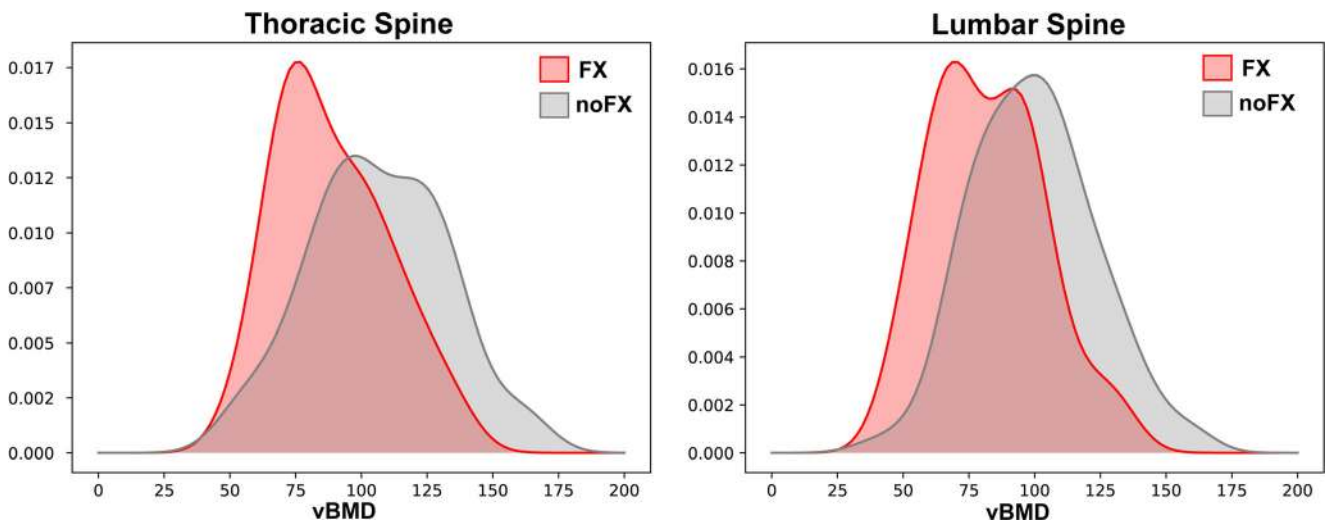
Fig. 2 The mean volumetric density distribution (vBMD) of the thoracic and lumbar spine in comparison with the FX and noFX group

Discussion

In this study, we developed a quantitative, automatic method based on opportunistic CT data to differentiate between patients with and without osteoporotic vertebral fractures. The results furnish evidence that regional vBMD and 3D texture analysis can discriminate between patients with and without vertebral fractures, without using data of fractured vertebrae. Parameters of all vertebral levels significantly contributed to this differentiation. Importantly, a combination of global and local BMD as well as 3D texture parameters outperformed volumetric BMD alone.

The possibility of opportunistic osteoporosis screening by assessing BMD in non-dedicated CT scans has widely been demonstrated [6, 10, 11]. Plenty of non-dedicated CT scans exist for this purpose, but widely vary regarding their acquisition and image reconstruction protocols. It has recently been pointed out that simple absorption measurements in Hounsfield units (HU) vary largely (up to 70 HU for the European Spine Phantom ESP 139) among scanners of different vendors, mainly due to different image reconstruction algorithms and radiation tubes with different voltage spectra [6]. Thus, HU values of different scanners or protocols should be converted to BMD values. For this purpose, two major methods have been proposed, namely phantomless (internal tissue calibration) [29] and phantom-based (either synchronous or asynchronous) [30–32] 1density calibrations, which can compensate for such systematic variations. In this study, we choose a direct, phantom-based calibration for the HU-BMD conversion. All of the scans used in this study were performed with intravenous contrast media. The effect of contrast-enhanced CT on vBMD has been studied for different settings, and linear conversion equations successfully corrected the systematic bias of this density variation [13]. In this study, we also corrected all vBMD values for contrast

media application. With such calibrations, a correct vBMD was calculated that is comparable among different studies and standard ACR thresholds for osteoporosis (i.e., ≤ 80 mg/ccm in the lumbar spine) apply.

In dedicated BMD measurements, the complete spine usually is considered one single skeletal site. However, a large variation of bone density and quality was demonstrated between different vertebral levels in elderly patients [33]. In an opportunistic screening approach, features from the complete thoracolumbar spine can be included to account for this variation. According to our analysis using the Gini index (i.e., importance of each feature for the classification), we demonstrated that every level of the spine was important. This suggests that as many vertebrae as possible should be included for an optimal prediction of the individual fracture risk.

The key for reliable prediction (i.e., classification) of fracture risk is the combination of BMD and other features of the vertebrae [5]. Recently, also advanced methods like finite element analysis have been applied in an opportunistic screening setting [10]. With FEA, biomechanical properties can be extracted to assess the bone strength by simulating loading conditions seen in daily lives [34]. The AUC for fracture prediction by vertebral strength was above 0.8 in a dedicated scenario [5]. However, the method is computationally intense and thus studies usually limited their evaluation on the lumbar region. It is also dependent on spatial resolution, scanning, and reconstruction parameters; consequently different acquisition parameters and scanners can lead to changes in FEA results [35].

Another technique to analyze bone properties is texture analysis. It can extract features complementary to vBMD by characterizing the distribution of voxel intensities. It is a well-established technique that can quantify regional variations on a global and local level [15]. 2D texture analysis is already used clinically for fracture risk assessment on existing DEXA
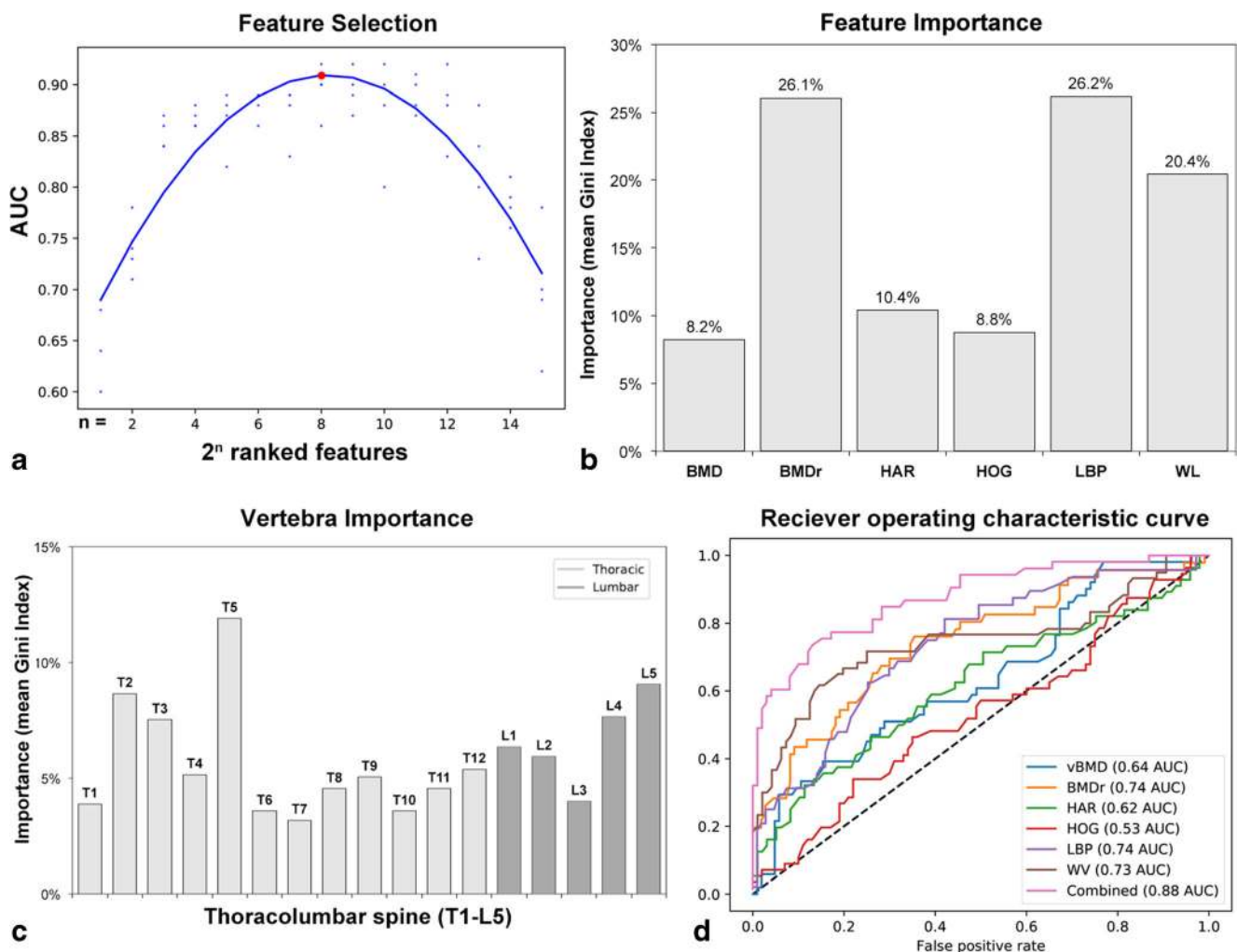
Fig. 3 Feature selection and importance. **a** Classification performance using feature selection. The ranked features according to the Gini importance (GI) are selected a in a 2^n fashion (i.e., 2, 4, 8, … .32768). The performance (AUC) of a fourfold cross-validation has been plotted for the increasing amount of selected features. The vertex (red dot) is used as the optimal cut of the fitted quadratic function (i.e., parabola) representing the overall performance of 0.88 AUC. **b** Composition of the set of important features. The mean Gini importance for each feature class of density and texture features is reported. Density features are split into global (vertebral level (vBMD)) and local features (sub-region level (BMDr)). **c** Composition of the set of important vertebrae. The mean Gini importance for each vertebra level is reported. **d** Comparison of the receiver operating characteristic (ROC) curves of each individual feature class and with the selected combined features.

images and named "trabecular bone score" (TBS) [36]. Our method fully exploits the 3D nature of the underlying CT datasets of the thoracolumbar spine. A 3D texture analysis was successfully used ex vivo in micro CT [37] and in vivo at the distal radius in high-resolution peripheral quantitative CT (HR-pQCT) images to describe the different facets of bone microarchitecture (texture patterns) in patients with and without fractures (classification performance; 0.67) [38] and after lung transplantation [39]. Like classical parameters of trabecular bone morphometry, LBPs describe distinct "patterns" of a texture and thus can discriminate between, e.g., plate-like and rod-like structures, but have the advantage of not being threshold-dependent. We demonstrated that clustered LBPs have the best individual classification performance, which make them quite robust and descriptive for opportunistic CT data, despite

the variation in image quality (i.e., contrast enhancement, noise, and spatial resolution) [23].

Recently, a number of machine learning approaches have been developed, which are able to work with very high-dimensional representations by unifying the feature selection and supervising learning tasks [26]. Handling the "small N large p" problem was one of the key features choosing the random forest (RF) over other classifiers. This means in our case it can handle few patients (i.e., few samples) with many features. A recent benchmark study also showed that it outperformed logistic regression [40]. Random forests are not only used for prediction but can also assess feature importance. The selection of informative features in the training set, a.k.a. feature selection, is a keystone in machine learning. Feature reduction is important to reduce overfitting of the

Table 2 Individual classification performance of each individual feature class (density and texture feature) using random forest (RF) classifier

	AUC	Specificity	Sensitivity
vBMD	0.64*	0.54	0.57
BMDr	0.74*	0.70	0.69
HAR	0.62*	0.59	0.59
HOG	0.53*	0.51	0.52
LBP	0.74*	0.68	0.71
WL	0.73*	0.68	0.69
Combined	0.88	0.78	0.77

*Statistical difference in AUC ($p < 0.01$) in comparison to combined features

results. If all possible features are used in the input space (i.e., feature vector), the results are compromised with unimportant, redundant features. This was evident also in our results, where AUC values decreased, if more than 256 features were included.

There are several limitations due to the retrospective nature of this study. First, we did not include any clinical scores such as age, BMI, or other parameters of the FRAX system [41] that might improve the performance; however, this was in particular relevant when using areal BMD derived by DXA [42]. Unfortunately, DXA was not available in this "opportunistic" MDCT dataset; thus, a comparison with conventional screening methods as DXA and FRAX was not possible. We also limited our analyses to vBMD and texture parameters, excluding cortical- or FEA-based biomechanical parameters. An inclusion of such parameters may further improve results [5, 43, 44]. However, we believe that the main findings of this study (feasibility of opportunistic screening using texture analysis; importance of all studied vertebra) are still valid without this information.

Second, we used a cross-validation instead of completely independent training and test sets due to the limited number of patients. However, in each of the four cross-validation datasets, results are only calculated for the test cases not
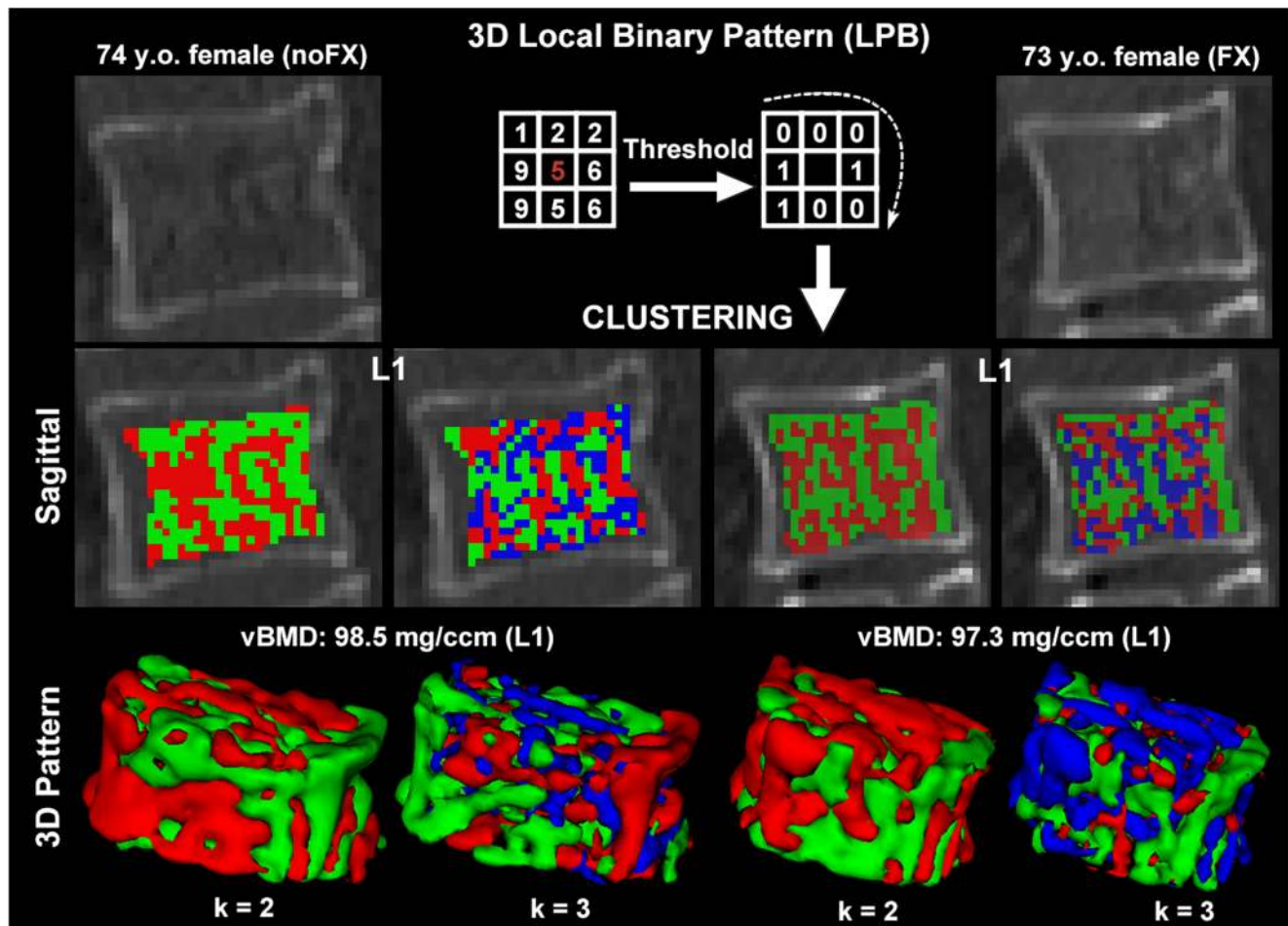

Fig. 4 Texture analysis using 3D local binary pattern (3D LBP). The procedure comprised the read-out of the intensity values around a circle centered on the pixel of interest in a binary fashion. If the surrounding pixel value is bigger than the central pixel, it gets the value of 1 and otherwise 0. Then clustering is used on the feature vector. Representatives in visualizing the differences in local binary patterns of L1 using 2 and 3 clusters (k) between a healthy 74-year-old female (noFX) and 73-year-old female from the fracture cohort (FX)

included in the respective training set. Additionally, the four-fold cross-validation was repeated ten times with stable results indicating that this might generalize to larger numbers of cases [45]. Third, we only separated patients with and without vertebral fractures. A prospective approach, predicting incident fractures, should be the aim of further studies.

In conclusion, the presented model based on a random forest classifier using 3D texture features in combination with trabecular bone mineral density features showed high potential for identifying patients with low bone quality susceptible to vertebral fractures in an opportunistic screening for osteoporosis. Parameters of all vertebral levels significantly contributed to this classification. Importantly, a combination of global and local BMD as well as 3D texture parameters outperformed volumetric BMD alone.

Compliance with ethical standards Ethics approval was obtained from the local ethics committee (11/5022A1). Due to the retrospective nature of the study, the need for informed consent was waived.

Conflict of interest Alexander Valentinitsch, Stefano Trebeschi, Johannes Kaesmacher, Cristian Lorenz, and Thomas Baum declare that they have no conflicts of interest. JSK reports research grants from the DFG, ERC, and Nivida Corporation related to the manuscript, as well as travel support from Kaneka Europe and Speaker Honorarium from Philips Healthcare, not related.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Center JR, Nguyen TV, Schneider D, Sambrook PN, Eisman JA (1999) Mortality after all major types of osteoporotic fracture in men and women: an observational study. Lancet 353:878–882. https://doi.org/10.1016/S0140-6736(98)09075-8
2. Kaesmacher J, Schweizer C, Valentinitsch A, Baum T, Rienmüller A, Meyer B, Kirschke JS, Ryang YM (2017) Osteoporosis is the most important risk factor for odontoid fractures in the elderly. J Bone Miner Res 32:1582–1588. https://doi.org/10.1002/jbmr.3120
3. Kanis JA, McCloskey EV, Johansson H, Cooper C, Rizzoli R, Reginster J-Y et al (2013) European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 24:23–57. https://doi.org/10.1007/s00198-012-2074-y
4. Cummings SR, Bates D, Black DM (2002) Clinical use of bone densitometry: scientific review. JAMA 288:1889–1897
5. Wang X, Sanyal A, Cawthon PM, Palermo L, Jekir M, Christensen J, Ensrud KE, Cummings SR, Orwoll E, Black DM, for the Osteoporotic Fractures in Men (MrOS) Research Group, Keaveny TM (2012) Prediction of new clinical vertebral fractures in elderly men using finite element analysis of CT scans. J Bone Miner Res 27:808–816. https://doi.org/10.1002/jbmr.1539
6. Engelke K (2017) Quantitative computed tomography-current status and new developments. J Clin Densitom 20:309–321. https://doi.org/10.1016/j.jocd.2017.06.017
7. Melton LJ, Riggs BL, Keaveny TM, Achenbach SJ, Kopperdahl D, Camp JJ et al (2010) Relation of vertebral deformities to bone density, structure, and strength. J Bone Miner Res 25:1922–1930. https://doi.org/10.1002/jbmr.150
8. Issever AS, Link TM, Kentenich M, Rogalla P, Schwieger K, Huber MB, Burghardt AJ, Majumdar S, Diederichs G (2009) Trabecular bone structure analysis in the osteoporotic spine using a clinical in vivo setup for 64-slice MDCT imaging: comparison to microCT imaging and microFE modeling. J Bone Miner Res 24:1628–1637. https://doi.org/10.1359/jbmr.090311
9. Damilakis J, Adams JE, Guglielmi G, Link TM (2010) Radiation exposure in X-ray-based imaging techniques used in osteoporosis. Eur Radiol 20:2707–2714. https://doi.org/10.1007/s00330-010-1845-0
10. Schwaiger BJ, Kopperdahl DL, Nardo L, Facchetti L, Gersing AS, Neumann J, Lee KJ, Keaveny TM, Link TM (2017) Vertebral and femoral bone mineral density and bone strength in prostate cancer patients assessed in phantomless PET/CT examinations. Bone 101:62–69. https://doi.org/10.1016/j.bone.2017.04.008
11. Lee DC, Hoffmann PF, Kopperdahl DL, Keaveny TM (2017) Phantomless calibration of CT scans for measurement of BMD and bone strength-inter-operator reanalysis precision. Bone 103:325–333. https://doi.org/10.1016/j.bone.2017.07.029
12. Genant HK, Wu CY, van Kuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 8:1137–1148. https://doi.org/10.1002/jbmr.5650080915
13. Kaesmacher J, Liebl H, Baum T, Kirschke JS (2017) Bone mineral density estimations from routine multidetector computed tomography: a comparative study of contrast and calibration effects. J Comput Assist Tomogr 41:217–223. https://doi.org/10.1097/RCT.0000000000000518
14. Hussein AI, Jackman TM, Morgan SR, Barest GD, Morgan EF (2013) The intravertebral distribution of bone density: correspondence to intervertebral disc health and implications for vertebral strength. Osteoporos Int 24:3021–3030. https://doi.org/10.1007/s00198-013-2417-3
15. Valentinitsch A, Trebeschi S, Alarcón E, Baum T, Kaesmacher J, Zimmer C, Lorenz C, Kirschke JS (2017) Regional analysis of age-related local bone loss in the spine of a healthy population using 3D voxel-based modeling. Bone 103:233–240. https://doi.org/10.1016/j.bone.2017.06.013
16. Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C (2009) Automated model-based vertebra detection, identification, and segmentation in CT images. Med Image Anal 13:471–482. https://doi.org/10.1016/j.media.2009.02.004
17. Klein S, Pluim JPW, Staring M, Viergever MA (2008) Adaptive stochastic gradient descent optimisation for image registration. Int J Comput Vis 81:227–239. https://doi.org/10.1007/s11263-008-0168-y
18. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern Syst 3:610–621. https://doi.org/10.1109/TSMC.1973.4309314
19. Albregtsen F (1995) Statistical texture measures computed from gray level coocurrence matrices. Image Processing Laboratory

20. Klaser A, Marszalek M, Schmid C. A Spatio-Temporal Descriptor Based on 3D-Gradients. 2008

21. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971–987. https://doi.org/10.1109/TPAMI.2002.1017623

22. Fehr J, Burkhardt H. 3D rotation invariant local binary patterns.2008. pp. 1–4. doi:https://doi.org/10.1109/ICPR.2008.4761098

23. Leung T, Malik J (2001) Representing and recognizing the visual appearance of materials using three-dimensional Textons | SpringerLink. Int J Comput Vis 43:29–44

24. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation- IEEE Xplore Document. IEEE Trans Pattern Anal Machine Intell 11:674–693

25. Akbarizadeh G (2012) A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images- IEEE Xplore Document. IEEE Trans Geosci Remote Sens 50:4358–4368

26. Breiman L (2001) Random forests. Mach Learn 45

27. Verikas A, Gelzinis A, Bacauskiene M (2011) Mining data with random forests: a survey and results of new tests. Pattern Recogn 44:330–349

28. Statnikov A, Wang L, Aliferis CF (2008) A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 9:319. https://doi.org/10.1186/1471-2105-9-319

29. Mueller DK, Kutscherenko A, Bartel H, Vlassenbroek A, Ourednicek P, Erckenbrecht J (2011) Phantom-less QCT BMD system as screening tool for osteoporosis without additional radiation. Eur J Radiol 79:375–381. https://doi.org/10.1016/j.ejrad.2010.02.008

30. Engelke K, Libanati C, Liu Y, Wang H, Austin M, Fuerst T, Stampa B, Timm W, Genant HK (2009) Quantitative computed tomography (QCT) of the forearm using general purpose spiral whole-body CT scanners: accuracy, precision and comparison with dual-energy X-ray absorptiometry (DXA). Bone 45:110–118. https://doi.org/10.1016/j.bone.2009.03.669

31. Mao SS, Li D, Luo Y, Syed YS, Budoff MJ (2016) Application of quantitative computed tomography for assessment of trabecular bone mineral density, microarchitecture and mechanical property. Clin Imaging 40:330–338. https://doi.org/10.1016/j.clinimag.2015.09.016

32. Brown JK, Timm W, Bodeen G, Chason A, Perry M, Vernacchia F, DeJournett R (2016) Asynchronously calibrated quantitative bone densitometry. J Clin Densitom 20:216–225. https://doi.org/10.1016/j.jocd.2015.11.001

33. Eckstein F, Lochmüller E-M, Lill CA, Kuhn V, Schneider E, Delling G, Müller R (2002) Bone strength at clinically relevant sites displays substantial heterogeneity and is best predicted from site-specific bone densitometry. J Bone Miner Res 17:162–171. https://doi.org/10.1359/jbmr.2002.17.1.162

34. Matsumoto T, Ohnishi I, Bessho M, Imai K, Ohashi S, Nakamura K (2009) Prediction of vertebral strength under loading conditions occurring in activities of daily living using a computed tomography-based nonlinear finite element method. Spine 34:1464–1469. https://doi.org/10.1097/BRS.0b013e3181a55636

35. Dragomir-Daescu D, Salas C, Uthamaraj S, Rossman T (2015) Quantitative computed tomography-based finite element analysis predictions of femoral strength and stiffness depend on computed tomography settings. J Biomech 48:153–161. https://doi.org/10.1016/j.jbiomech.2014.09.016

36. McCloskey EV, Odén A, Harvey NC, Leslie WD, Hans D, Johansson H et al (2016) A meta-analysis of trabecular bone score in fracture risk prediction and its relationship to FRAX. J Bone Miner Res 31:940–948. https://doi.org/10.1002/jbmr.2734

37. Räth C, Monetti R, Bauer J, Sidorenko I, Müller D, Matsuura M, Lochmüller EM, Zysset P, Eckstein F (2008) Strength through structure: visualization and local assessment of the trabecular bone structure. New J Phys 10:125010. https://doi.org/10.1088/1367-2630/10/12/125010

38. Valentinitsch A, Patsch JM, Burghardt AJ, Link TM, Majumdar S, Fischer L, Schueller-Weidekamm C, Resch H, Kainberger F, Langs G (2013) Computational identification and quantification of trabecular microarchitecture classes by 3-D texture analysis-based clustering. Bone 54:133–140. https://doi.org/10.1016/j.bone.2012.12.047

39. Fischer L, Valentinitsch A, DiFranco MD, Schueller-Weidekamm C, Kienzl D, Resch H et al (2015) High-resolution peripheral quantitative CT imaging: cortical porosity, poor trabecular bone microarchitecture, and low bone strength in lung transplant recipients. Radiology 274:473–481. https://doi.org/10.1148/radiol.14140201

40. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. 2017

41. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E (2008) FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int 19:385–397. https://doi.org/10.1007/s00198-007-0543-5

42. Donaldson MG, Palermo L, Schousboe JT, Ensrud KE, Hochberg MC, Cummings SR (2009) FRAX and risk of vertebral fractures: the fracture intervention trial. J Bone Miner Res 24:1793–1799. https://doi.org/10.1359/jbmr.090511

43. Bouxsein ML, Melton LJ, Riggs BL, Muller J, Atkinson EJ, Oberg AL et al (2006) Age- and sex-specific differences in the factor of risk for vertebral fracture: a population-based study using QCT. J Bone Miner Res 21:1475–1482. https://doi.org/10.1359/jbmr.060606

44. Hussein AI, Morgan EF (2013) The effect of intravertebral heterogeneity in microstructure on vertebral strength and failure patterns. Osteoporos Int 24:979–989. https://doi.org/10.1007/s00198-012-2039-1

45. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley-Interscience