**Texture Analysis Based on Gabor filters Improves the Estimate of Bone Fracture Risk from DXA Images**

Rui-Sheng Lua,\*, Elaine Dennisonb, Hayley Denisonb, Cyrus Cooperb, Mark Taylora, and Murk J. Bottemaa

a School of Computer Science, Engineering and Mathematics, Flinders University, Adelaide, SA, Australia

b MRC Lifecourse Epidemiology Unit, Faculty of Medicine, University of Southampton, Southampton, UK

**\*Corresponding author:** Rui-Sheng Lu

School of Computer Science, Engineering and Mathematics, Flinders University

1284 South Road (Tonsley), Clovelly Park, SA 5042, Australia

T: +61 8 8201 3652 | Fax: +61 8 8201 2904

Email: lu0141@flinders.edu.au

**Texture Analysis Based on Gabor filters Improves the Estimate of Bone Fracture Risk from DXA Images**

**Abstract**

We investigate whether novel parameters derived from a DXA scan might improve fracture risk estimation. In this study, we analysed hip DXA scans from 29 older adults with a history of fragility fracture and 90 non-fractured controls. Active shape models (ASM) and active appearance models (AAM) (Cootes et al. 1995; Cootes et al. 2001) were used to allow a quantitative characterization of the shape and gross structure of the proximal femur. We also performed image texture analysis applied to various regions of interest (ROI). Feature selection was used to determine which method, or combination of methods was best to discriminate between the fracture and control groups. Texture features derived from Gabor filters (Petkov 1995) in combination with total T-score provided better estimates of risk (AUC =0.787) than the standard measures of areal bone mineral density (aBMD) or total T-score alone (AUC =0.699 and 0.692, respectively). Estimates of risk were more accurate when the texture was measured on the whole femoral neck compared to other regions. The features extracted from the active models were weaker with poor classification performance (AUC <0.570). This study shows that image texture based on Gabor filters can complement the standard measures to improve fracture risk estimation.

**Keywords** active shape and appearance models; bone fracture; Gabor filters; textons; texture; DXA

**Introduction**

Many older people, especially women are at a significant risk of suffering fracture (Cummings et al. 1993; Stone 2003). Calculation of aBMD by DXA is often an important step in the pathway to estimation of fracture risk, as a 1SD decrease in bone mineral density is associated with an approximate doubling of fracture risk (Marshall et al. 1996). However, some studies have suggested that the ability of aBMD to predict fracture risk is limited (Marshall et al. 1996; Wilkin et al. 2001; Kanis 2002; Robbins et al. 2005; Pulkkinen et al. 2010). More than 50% of fractures occur among people without low aBMD, in part because low aBMD is not the only risk factor for fractures (Pulkkinen et al. 2010; Kanis 2002 ), and some risk factors for osteoporosis appear to act independently of BMD (Cummings et al. 1995).

Although aBMD alone may not provide an accurate prediction of fracture risk, other information available from DXA images such as the femoral geometry of bone may provide further information regarding the risk of fracture either on its own or in combination with aBMD (Pulkkinen et al. 2010). Researchers have explored the influence of geometric parameters on the risk of fracture of the femur. Pulkkinen et al. proposed that the aBMD T-score ≤−2.5 criterion discriminates the risk of trochanteric fractures, whereas geometric risk factors are able to discriminate cervical fracture cases from control cases with similar aBMD (Pulkkinen et al. 2010). Hip axial length, femoral length, femoral neck width, and neck shaft angle are the most examined geometric parameters with respect to fracture risk, but findings have been inconsistent (Dinçel et al. 2008). For instance, some studies observed that fracture risk is associated with hip axial length (Faulkner et al. 1994; Gnudi et al. 1999), but some did not (Alonso et al. 2000; Pande et al. 2000). Michelotti, et al. reported that fractured cases had thinner femoral cortices, larger femoral heads, and larger femoral neck diameters than controls (Michelotti and Clark 1999). Dinçel, et al. observed that there was a significant increase in the ratio between femoral neck width and femoral length in the fracture group (Dinçel et al. 2008). A wider femoral neck and shaft, and a bigger neck-shaft angle were observed by Gregory et al. where both male and female fracture subjects were compared with controls (Gregory and Aspden 2008). Some of these geometrical parameters are highly correlated (Gnudi et al. 1999). Some studies have observed that femur bone dimensions continue to change with age so that it is difficult to fit these geometrical parameters to individual DXA images (Heaney et al. 1997; Beck et al. 2000). However, little has been done to explore whether additional information collected as part of a femoral DXA scan improves the prediction rate for all fractures, not just of the femur.

Rather than using discrete measures of geometry, active shape models (ASM) may be used to capture all of the geometric variation in femoral shape in the population of samples. The image object in ASM is represented by a number of principal modes that describe the variation in the shape, and therefore each is an independent descriptor of the shape for an object (Cootes et al. 1995). These distinct descriptors from ASM have been used to describe morphometric features to identify subjects at high risk of hip fracture (Gregory et al. 2007). Gregory, et al. demonstrated that using ASM to predict the risk of hip fracture is more effective than aBMD and discrete geometry measurements. The study also showed that the combination of Ward’s triangle aBMD and ASM can be used to improve the accuracy to 90% (Gregory et al. 2004). While ASM describes the shape of the proximal femur, it does not provide any information regarding variation of bone density within the femur. Active appearance models (AAM), on the other hand, match a statistical model to the shape and density distribution of the proximal femur (Goodyear et al. 2013; Bredbenner et al. 2014). ASM and AAM have been shown to be as accurate as manual observers in the measurement or location of vertebral shape(Smyth et al. 1997), hip fractures(Gregory et al. 2005; Lynch et al. 2009; Baker-Lepain et al. 2011; Sarkalkan et al. 2014), and other medical imaging applications (Cootes et al. 1994; Cootes and Taylor 2001; Ginneken et al. 2002).

Furthermore, recent work has demonstrated that decreased bone strength is not only dependent on low aBMD but also on trabecular bone microarchitecture (Lespessailles et al. 2006). Several texture analysis methods have been developed to characterise the change of trabecular patterns or textures on proximal femur. Yap, et al. proposed an adaptive sampling method such that the sampled locations in different images correspond to consistent locations, where a set of Gabor filters were applied to each sampled region to extract the texture features. He suggested that this method further improved the overall performance of fracture detection especially when combined with neck-shaft angle measurement (Yap et al. 2004). Lim, et al. included neck-shaft angle measurements, Gabor filters, Markov random field texture, and intensity gradient and concluded that fracture detection rate was improved significantly by combining these methods (Lim et al. 2004). Similarly, Pramudito, et al. used three different texture analysis methods including Gabor filters, wavelet transforms and fractal dimension to extract features that represent the structural change in trabecular pattern, and compared these methods with the corresponding Singh index grading system (Pramudito et al. 2007). The features extracted using Gabor filters were significantly correlated with Singh index.

Julesz was the first to propose the term “textons” and stated that textons are the putative units of pre-attentive human texture perception (Bela. 1981). In image analysis applications, local texture descriptors are mapped to a feature space, the dimension of which is the number of local texture descriptors used. Clusters in this feature space are called textons and represent commonly occurring local texture patterns. Each pixel in the image is then mapped to the textons closest to it in the feature space representation of the pixel. The image is represented by the histogram of textons occurrences (Varma et al. 2005). Textons-based approaches are simple to implement and often achieve good performance for texture image categorization and segmentation (Malik et al. 2001; Shotton et al. 2008). For medical imaging application, Petroudi et al. used textons-based features generated from a filter bank to classify mammograms into the four breast imaging reporting and data system (BI-RADS) classes (Petroudi et al. 2003). More recently, T. Jiang et al. proposed a textons-based classification system based on raw pixel representation along with a support vector machine (SVM) with radial basis function (RBF) kernels for the classification of emphysema in computed tomography images of the lung. Classification accuracy improved to 96.43% (Gangeh et al. 2010).

In this study, five methods - ASM, AAM, Gabor filters, Gabor filters-based textons, and local 3x3 neighbourhoods-based textons - derived from traditional DXA scanning are compared individually and in combination to assess how well they relate to fracture status in older Caucasian adults. The methods considered include standard measures aBMD, T-score, statistical methods ASM and AAM as well as two classes of texture methods based on textons.

The object of this study is to develop a method for predicting the risk of fragility fracture generally - not just hip fracture - based detailed analysis of DXA images of the hip. A fragility fracture or called low-energy fracture is defined as a fracture resulting from minimal trauma falling from standing height or less, rather than any other type of trauma such as motor vehicle accident (Cooper et al. 1996). Predicting fragility fractures generally is more useful than predicting hip fracture alone as this avoids the necessity for subject undergoing an array of procedures to assess risk various fracture types separately.

**Materials and methods**

***Data and Software***

DXA images were prospectively obtained from 119 subjects, forming a subset of individuals participating in the Hertfordshire cohort study. Details of this study have been described previously (Syddall et al. 2005). In brief, DXA images where taken around the individuals’ 66th birthday (range 59–74 years old), and fracture status determined by a lifestyle questionnaire. All subjects underwent a scan of the unfractured left hip by DXA using a Hologic QDR4500 scanner. The pixel resolution of the digitized proximal femur radiographs was 250 x 300 pixels and effective depth of 8 bits. The fracture group consisted of 29 white subjects with reported fragility fractures (10 males, 19 females), and the control group consisted of 90 white subjects without a history of fragility fractures (33 males, 57 females). The types of fractures were not restricted to hip fracture but included wrist, hip, lower limb, and spine fractures. For each subject, aBMD for the femoral neck and total hip, T-score for the femoral neck and total hip, hip axis length (HAL), neck shaft angle (NSA) were recorded. 60 subjects (15 fracture subjects, 45 control subjects) were randomly selected to form the training set, and 59 subjects (14 fracture subjects, 45 control subjects) were reserved as a testing set.

All computations were performed using in-house written software using the Matlab® programming platform.

***Regions of Interest for Texture Analysis***

Various types of fractures can occur, ranging from minor inconveniences to severe, life-threatening fractures. The ROI for the proximal femur are typically the total proximal femur, femur neck, intertrochanteric region, trochanter, Ward’s triangle, and femoral shaft (Duboeuf et al. 1991; Marco et al. 2004; Vijay et al. 2011; Yujia et al. 2015). Evidence suggested that the femoral neck or total proximal femur in particular are the optimum sites for predicting the risk of fragility fractures (Cummings et al. 1993; Stone et al. 2003). Accordingly, texture features were computed on seven ROI to determine which provided the best indicators of risk. The ROI considered for texture analysis were the whole hip, the whole femoral neck, neck, ward’s triangle, narrow neck, inter- trochanter and femoral shaft (Figure. 1).

 [Figure 1 near here].

***Active shape model (ASM) and Active appearance model (AAM)***

This section provides a detailed description of ASM and AAM. In principle, these methods apply to data of any dimension but since this study concerns DXA images, the formulation here is described for two-dimensional (2-D) arrays.

The ASM and AAM of the proximal femur were created from the training set of DXA images based on the methods of Cootes and colleagues (Cootes et al. 1995; Cootes et al. 2001). The first step was to manually identify enough points on the boundary of the femur in each image in the training set to accurately capture shape information. To place landmark points along the boundary, the digital DXA image was displayed on a computer screen and one of the authors (R.S. Lu) used mouse clicks to identify boundary points. Mouse clicks were captured and to record point locations as (*x*, *y*) coordinates. In this study, 44 landmark points were used to outline the proximal femur in each image. Care was taken to place the landmark points consistently across all the images. To do this, closely spaced landmark points were placed along significant morphological features of the femur, including the femoral neck and greater trochanter (Fig. 2). The lessor trochanter was not included as part of the femur outline because its appearance is highly variable in DXA images and the boundary is often not very clear. The appearance of the lesser trochanter depends heavily on the orientation. The same number of landmark points were used across all the images for a particular signiﬁcant feature. The remaining landmark points were placed roughly equally spaced between the signiﬁcant features (Fig. 2). Note that the landmark points were deliberately spaced non-uniformly along the boundary in order to concentrate shape information at highly curved parts of the femur outline. The literature reports that the exact locations of the landmark points does not signiﬁcantly impact ﬁnal results as long as care is taken to be somewhat consistent between subjects (Kim and Sung 2009). One reason for the robustness of these models against the exact placement of landmark points is that in the next step, the method of principal component analysis (PCA) removes ﬁne shape anomalies while retaining signiﬁcant shape information.

[Figure 2 near here].

In order to compare shapes, the landmark points in the training images were aligned automatically into a common coordinate frame by using scaling, rotation and translation to eliminate differences between femur size, orientation and position within the image (Cootes et al. 1995). The coordinates of the aligned landmark points for the femur in image *i* are denoted by (*xi*,1, *yi*,1), (*xi*,2, *yi*,2), ..., (*xi*,44, *yi*,44). These were stacked to form the shape vector *Xi* for the femur in image *i* as

*Xi* = (*xi*,1, *yi*,1, *xi*,2, *yi*,2, ..., *xi*,44, *yi*,44)*T* , (1)

where the superscript *T* denotes the transpose of the vector. Note that the vector *Xi* has length 88.

For the collection of shape vectors from the training set, {*Xi*, *i* = 1, 2, ..., *N*}, the mean vector is

 , (2)

and the covariance matrix is the 88 × 88 matrix

. (3)

PCA consists of ﬁnding the eigenvalues and associated normalised eigenvectors Ф*s* of *s*, listed in order of decreasing eigenvalue. The eigenvalues indicate the proportion of the information content encoded in the corresponding eigenvector. PCA is a general method but when applied in this context, the eigenvectors are called principal modes. The principal modes are mutually orthogonal (Ф*s*Ф*tT* = 0 if *s* ≠ *t*) and so PCA separates the information content of the vectors *Xi* into uncorrelated principal modes Ф*s*.

For a shape vector from the training set,

, (4)

where the coefficients *bs* are computed by

. (5)

For a shape vector *X* from a femur outside the training set

. (6)

The reason for using PCA is to remove redundancies in the shape information and simultaneously remove shape details that are insigniﬁcant, including small shape anomalies resulting from inconsistencies in placing boundary points. To do this, only principal modes corresponding to the top eigenvalues are retained to represent the shape vectors. In this study, the training data was used to determine that the top 12 principal modes represent 99 percent of the total information content and so these 12 modes were used to represent the shape of each femur in all training and testing images.

The ﬁnal vector representing the femur shape is denoted by  and was deﬁned as

. (7)

ASM models the shape of the femur only. ASM does not model the grey-scale variation within the femur. Two femurs with identical proﬁles in the DXA image may have very different risks of fracture due to the amount and distribution of bone tissue within the boundary of the femur in the image. Such information manifests in the grey-scale values of the pixels representing the femur. In order to explain shape and grey-scale variation simultaneously AAM, on the other hand, is generated by combining a model of shape variation as well as a model of grey-scale variation in a shape-normalised frame to explain shape and grey-scale simultaneously (referred to as the appearance). To access this information of the grey-scale variation, a procedure similar to the ASM was performed on the gray-scale values of pixels within the region of the femur in the DXA images. To do this, each training image was aligned automatically as described for ASM and used to generate generated the mean shape. Each training image was warped so that its boundary points match those of the mean shape, obtaining a shape-free patch. Then, the grey-scale values at 82,830 evenly spaced pixels from the shape-free image over the region covered by the mean shape were recorded (Fig. 3) (Cootes et al. 2001). Because shape-free patches were used, the locations of the pixels were reasonably consistent between femurs in different images.

[Figure 3 near here].

For image *i*, the numbers *gi*,1, *gi*,2,..., *gi*,*m* represent the grey-scale values at pixels 1,2, ..., *m* within the region of the femur. Here *m* = 82,830. The gray-scale vector associated with image *i* is *Gi* given by

*Gi* = (*gi*,1, *gi*,2,..., *gi*,*m*)T , (8)

and the average gray-scale vector for the training set is

. (9)

The same steps used for ASM were be used to construct a grey-scale model for the femur. The grey-scale model was combined with the shape into a single model. Since there may be correlation between the shape and grey-scale models, PCA was applied to the combined shape and grey-scale representations to generate the appearance model. In this study, the top 57 principal modes from the appearance model were found to represent 99 per cent of total appearance information and so only the top 57 principal modes were used in AAM to represent the appearance of each femur. Thus for any appearance vector (shape and gray-scale) from the training or testing set the appearance of the femur was represented by  and, and were deﬁned as

, (10a)

 , (10b)

where andare the principal modes of shape and gray-scale variation, respectively. The parameter *ca* controls the shape  as well as gray-scale , and thus the set *ca*, a = 1, 2, …, 57 was used as the set of appearance features for each subject.

***Gabor filters***

The Gabor features space consists of responses which are produced by convolving the original image with a bank of 2-D (real) Gabor kernel functions at several different scales (spatial frequencies) and orientations (Petkov 1995; Grigorescu 2002). The Gabor kernel function is the product of a Gaussian and a cosine function. Let denote an image and denote the Gabor kernel function with scale value and orientation value *θ*. The Gabor filter response at this scale and orientation to this image is expressed as

 (11a)

 (11b)



,

where *γ* = 0.5 is a constant, called the spatial aspect ratio that determines the ellipticity of the receptive field, *λ* represents the wavelength (so 1/*λ* is the spatial frequency),  represents the orientation, *σ* is the deviation and determines the size of the receptive field (the ratio σ/λ determines the spatial frequency bandwidth), , is the phase offset that determines the symmetry of  with respect to the origin; the phase offset values  and  correspond to symmetric functions (also called even), while *φ* = -π/2 and *φ* = π/2 correspond to anti-symmetric functions (or also called odd). The output of the Gabor filter is a number that indicates how much the image intensity pattern at position (*x*, *y*) in the image resembles a line segment at orientation theta and of width and length characterised by the other parameters. By recording these filter outputs at locations (*x*, *y*) within the femur (or a subregion of interest) a summary of oriented structure is obtained. This oriented structure is a measure texture. The terminology is standard in image analysis but it is important to note that this is an image texture resulting from patterns in tissue structure and is not a physical texture of the bone, for example.

In this paper the outputs of a symmetric (*φ* = 0) and an anti-symmetric filter (*φ* = -π/2) at each image pixel were combined into a single value called the Gabor energy (Petkov 1995; Grigorescu 2002). There are no general methods for the selection of Gabor filter parameters, which is often a vague and application dependent task. Here orientation and spatial frequency were sampled uniformly and fixed values of the remaining parameters were set according to the literature (Turner 1986; Fogel 1989; Kruizinga 1999; Petkov 1995; Grigorescu 2002). Thus, the value of the remaining parameters for Gabor filters constructed from the Gabor kernel function in Equation 11b were as follows. (1) Four spatial frequencies, 1/*λ* = 1/16, 1/18, 1/26 and 1/32, were used in the cosine factor of the Gabor kernel function (Turner 1986; Fogel 1989). (2) Six orientations, *θ = n*π*/6*, *n* = 0, 1, 2, 3, 4, 5, were used. (3) The spatial aspect ratio was set as its default value, *γ =* 0.5 (Petkov 1995; Grigorescu 2002). (4) The value of *σ* was specified by setting the half response spatial frequency bandwidth, *b=*1 (Kruizinga 1999; Grigorescu 2002)*.* The value of *b* is related to the ratio as follows

 . (12)

Thus 24 Gabor-filtered images, were produced for each original image (Fig. 4). For each Gabor image, the total Gabor energy for a particular choice of *λ* and *θ* was computed as

, (13)

where  is the magnitude of the Gabor-filtered image .

[Figure 4 near here].

***Textons***

Two classes of textons features were extracted from images in this study: Gabor filters based textons and local 3x3 neighbourhoods-based textons. To compute Gabor filters textons, the 24 Gabor filters described in previous section were applied to all the ROI in all the training images. This resulted in 24 filter responses for each pixel in each ROI. For every pixel p, the 24 responses at p were used to form a vector vp of length 24. For each single ROI type (for example for the whole femoral neck) the vectors vp from all the pixels in this ROI for all images in the training set were viewed as points in a 24 dimensional representation space. K-means clustering with K = 20 was used to identify 20 clusters in the representation space for each ROI type. These 20 clusters are the Gabor textons for each ROI type. Next each pixel in the ROI was identified with the textons closest to vp in the representation space for that ROI type. Each ROI was then represented by the normalised histogram of the textons occurrences in the ROI. Thus each ROI in each training image was represented by a vector of length 20 called the Gabor textons distribution.

The 3 x 3 textons representation was computed similarly. In this case, for each pixel p, the image intensities of the 8 neighbouring pixels were used to form a vector vp of length 8. The remaining steps followed those for constructing the Gabor textons representation. Thus the vectors vp for each ROI type were collected into an 8 dimensional representation space and K-means clustering with K = 20 was used to find 20 textons (Fig. 5). Pixels were identified with one of these textons according to the closest textons in the representation space to vp. Each ROI in each image was represented by a vector of length 20 called the neighbour textons distribution (Fig. 5).

[Figure 5 near here].

In the testing stage, the same procedure was repeated as in the training stage except that the clustering step was omitted. Thus, the vectors vp were computed for each pixel, but the pixel p was associated with the texton from the training step closest to vp in the representation space. The Gabor textons and the 3x3 neighbour textons were computed analogously to these quantities for the training images.

***The classification***

With 12 ASM features, 57 AAM features, 24 Gabor filters features, 20 Gabor texton features and 20 3x3 texton features, each ROI in each image was represented by a feature vector of length up to 133. Since classification based on this many features with only 60 training and 59 testing examples is unreliable, feature selection was used to reduce the dimensionality of the feature space. As a rule of thumb, the number of features should be approximately log10 (N) where N is the number of samples available. Accordingly, three features were seen as the most reasonable number of features to select. In addition, all individual features and all combinations of 6 features were considered for comparison. The method of exhaustive search was used three times for each ROI type to reduce the number of features for classification to 1, 3 and 6 features respectively. For each ROI type, the parameters for three Fisher linear classifiers were determined using the 1, 3 and 6 dimensional feature spaces. Only training data was used for the feature selection step and fitting the Fisher classifier. The classifiers trained in this way were then applied to the respective ROI types in the testing images.

The procedure described thus far provided estimates of how well the shape and appearance models and texture measures were able to predict fractures. Also of interest was the performance of combining these methods with aBMD and/or T-score. To assess the combined risk factors, the steps above were repeated but with standard risk factors included in the classification step. To form a comprehensive picture of the relative contributions from these methods, features from these methods were combined in two ways: (1) by including clinical standards (total aBMD and total T-score) in the pool of features prior to feature selection, and (2) by augmenting clinical standards to the set of shape, appearance and texture features after feature selection.

A useful measure of classification performance on a clinical test is the area under the receiver operating curve (AUC) ranging from 0.0 to 1.0 (Metz 1978). In general, the higher AUC score, the better classification performance. Thus the performance of classification for each method, and their combination were reported as the area under the receiver operating curve.

The ROI and feature combination yielding the best performance on the testing data was further evaluated using leave-one-out cross validation patch (LOOCV) (Hastie et al. 2009).

**Results**

***Baseline characteristics of the subjects***

There were no significant differences between the case and control group with regard to age, neck aBMD, neck T-score, HAL and NSA. As expected, the case group had lower total aBMD and total T-score compared to control group (*p* < 0.05, t-test) (Table 1).

 [Table 1 near here].

***Classification results using a single method***

The standard risk factors (total aBMD, neck aBMD, total T-score and Neck T-score) each consist of a single value, viewed here as a single feature. As a single feature, total aBMD outperformed the other single features including the best single features selected from ASM and AAM (Table 2a). Increasing the number of features for ASM and AAM improved the training scores but the testing scores were poor in each case (Table 2a).

[Table 2a near here].

For Gabor filters, neighbour textons and Gabor textons on individual ROI, the best single feature in terms of AUC for the testing data was a Gabor filter feature measured on the whole femoral neck region (Table 2b). For the best combinations of three features, the best performance on the testing data was provided by the method of Gabor textons applied to the femoral shaft (Table 2b). For the best six features the highest AUC score for testing data was 0.603 from Gabor filters on the femoral neck.

[Table 2b near here].

***Discriminant analysis using combinations of two methods***

We next sought to study the performance obtained by combining features from different methods. Two strategies for combining features were considered. The first strategy was to gather the features from a particular method and ROI type (for example, the 24 Gabor filters on the whole femoral neck) with a standard measure (for example total aBMD) to form a new initial set of features and then find the best combination of 1, 3 and 6 features from this new set (gather features from different methods, then select optimal combinations). The second strategy was to select optimal sets of features within one of the methods as in the previous section, but then augment this set with one of the standard features (select optimal combinations, then gather features from different methods). The same trends were obtained by both methods. Hence only the result of the first strategy is presented (Table 3).

The combination of total T-score and AAM demonstrated higher discriminative ability for fracture than other combinations at the training stage (AUC =0.907 in 6 features). The best predictive capacity at the testing stage was obtained when total T-score was combined with Gabor filters on the whole femoral neck (AUC =0.787 in 6 features) (Table 3).

[Table 3 near here]

***Classification performance using combinations of multiple methods***

Feature combinations from all methods were next considered. The number of such features over all five methods and all ROI was too large to conduct exhaustive-search feature selection and so the number of features was reduced by considering only the top six features from each method and restricting attention to the three best performing methods; ASM (6 features), ASM (6 features) and Gabor filters on the whole hip (6 features) plus the standard measures total aBMD and total T-score, forming a 20 dimensional feature vector. Exhaustive searching was used on the training set, and we found that a perfect training performance (AUC = 1.000) could be achieved with a combination of 11 features. In addition, all individual features were considered and all combinations of 6 features were considered for comparison. The resulting features and classifiers were used to estimate performance on unseen data using the testing set (Table 4).

[Table 4 near here].

***Principal modes and classification performance***

In comparing ASM and AAM, the variation in the data explained by the first six principal modes was much larger (more than 90 percent) for ASM than for AAM (less than 30 percent). However, the testing AUC for AAM using the first six principal modes was higher (AUC = 0.707) than the testing AUC score for ASM (AUC = 0.658) using the first six principal modes.

***Comparison between ASM and AAM and with previous studies***

Additional comparisons were made between ASM/AAM and previous studies (Gregory et al. 2004; Baker-Lepain et al. 2011; Goodyear et al. 2013). Comparisons with other studies were made in terms of study parameters (Table 5a), performance on training data (Table 5b) and performance on testing data (Table 5c).

[Table 5a near here].

[Table 5b near here].

[Table 5c near here].

***Leave-one-out cross validation***

Since the best results were obtained using T-score and a set of five Gabor filter features from the whole femoral neck, these features from this ROI form the recommended method for estimating fracture risk. Leave-one-out cross validation using these features resulted in an ROC curve similar to the ROC curve found using these features on the independent testing set (Figure 6). The leave-one-out cross validation AUC score was 0.762 compared to an AUC score of 0.787 for the independent testing set.

[Figure 6 near here].

**Discussion**

In this study texture information derived from Gabor filters in combination with total T-score provided a better estimate of fracture status than the standard measures of aBMD or total T-score alone. In particular, 6 features selected from Gabor filters computed over the whole femoral neck and total T-score gave an AUC more than 10 percent higher than aBMD alone, total T-score alone or combinations of textons features alone. The combination of Gabor filters and total T-score also outperformed risk estimates based on ASM and AAM, which, on their own, performed very poorly (Table 2a, Table 3). In addition, risk estimates based on the proximal femur were more reliable than estimates based on other regions of interest tested (Table 2b). These conclusions are based on estimates of risk obtained by testing the features and classifier parameters found during the training stage on an independent set of images. Leave-one-out analysis confirmed the predictive power of the T-score and the five Gabor filter features measured on the femoral neck for estimating fracture risk.

 Fracture cases in the data set included fractures of wrist, hip, lower limb, spine, etc. Accordingly, the results indicate that the methods presented here apply to predicting the risk of fracture generally. In many situations, predicting fracture generally is more useful than predicting only femur fracture. However, predicting femur fracture from DXA images of the femur is likely to be more accurate than predicting general fracture from the same DXA images. Unfortunately, sufficient data was not available in this study to measure femur fracture risk prediction.

 This study was limited by the quantity and quality of DXA images available. DXA images are widely available, but a study of the type reported here requires DXA images of fracture and matched non-fracture subjects. Only a few data sets exist satisfying this requirement and unfortunately, all such image data sets known to the authors are stored in proprietary image formats that prevent the application of novel image analysis methods. Accordingly, this study was based on relatively few images of low quality. While this limits the ability to properly estimate the full potential of the methods presented for estimating fracture risk, the application of these methods to higher resolution images can only improve results. Since improvement over existing methods was demonstrated on low quality images, this study suggests that estimates of fracture risk better than those reported here could be expected in practice using full resolution DXA images.

 This study included elderly white subjects only, and thus the results in this paper do not automatically extend to younger people or other ethnic groups. However, if image texture characteristics differ in other groups then the specific weighting and parameters reported here may need to be recomputed, but the methods presented here provide the framework for doing so. Finally, the individuals recruited were selected because they had been born in Hertfordshire, and continued to live there at the age of 60-75 years. However, we have previously demonstrated that the Hertfordshire populations studied have similar smoking characteristics and bone density to national figures (Syddall et al. 2005), suggesting that selection bias is minimal.

We compared our findings to three other studies (Gregory et al. 2004; Baker-Lepain et al. 2011; Goodyear et al. 2013). Differences must be interpreted with care due to disparity in the type of fractures considered, the landmarks used and the regions over which aBMD was computed. In this study, measurements taken from DXA images of the femur were used to estimate the risk of all fractures (femur, vertebrae and wrist), not just of the femur, whereas other studies have specifically focussed on the risk of fracture of the femur. Nevertheless, several observations may be made. First, the results for ASM alone are consistent over all the studies except the one that included the pelvis in the model (Goodyear et al. 2013). This may be due to the large number of initial features extracted in the study by Goodyear et al. Too many features may dilute the class-relevant shape information necessitating many more training examples. This may also explain the superior results in this study compared to previous use of AAM (Goodyear et al. 2013), but the discrepancy may also be due to the final number of features used in classification (Table 5a). Second, total aBMD was consistent over the studies for which this information was available and this supports the tacit assumption that the images used in all these studies are comparable in quality. In a study performed by Gregory et al, better prediction of hip fractures was achieved using neck and Ward aBMD (Gregory et al. 2004). Improved performance for predicting all fractures using neck aBMD was not supported in this study and Ward aBMD was not available. Third, in all studies, combining a form of aBMD with ASM resulted in better performance. Total aBMD alone was about the same for study by Goodyear et al (Goodyear et al. 2013) and the present study, but, as noted above, the performance of ASM was much lower in the study by Goodyear et al (Goodyear et al. 2013) than in the present study. However, the combination of ASM and total aBMD was the same in these studies. These results indicate that, if ASM is used effectively (pelvis not included, for example), then total aBMD does not contribute extra information, but if ASM is used less efficiently, then aBMD may compensate. Fourth, these studies do not necessarily support the notion that increasing the number of features improves performance. Notably, combining AAM, ASM and total aBMD in the study by Goodyear et al led to poorer results than combining just ASM and total aBMD (Goodyear et al. 2013) (Table 5).

The comparisons we made with other studies were based on training results instead of testing results (Table 5b). This was necessary because only training scores were presented in the papers cited. In the study performed by Baker-Lepain et al, a bootstrap procedure was used to determine that effects were different from what could be expected by chance alone if there was no actual effect, but that does not address the robustness of the reported performance scores (Baker-Lepain et al. 2011). The parameters determined during the development stage were not tested on an independent data set. In the study performed by Gregory et al, models were trained and tested on independent data but feature selection and classification was only performed on the resulting best model (Gregory et al. 2004). Thus, the selection of the best model was validated but feature selection and classification were not tested on an independent data set, and so results must be viewed as training results. Classification results based on training only may not be reliable, especially if feature selection is used in the training process. This is well documented in the literature (Hastie et al. 2009) and there were many examples in this study. By selecting six neighbour textons features on the whole hip ROI, an AUC score of 0.873 was found during the training stage but when the resulting classifier was applied to the set of test images, the AUC score reduced to 0.398 (Table 2b). Perfect classification was achieved by combining 11 ASM, AAM and Gabor filters features during training but performance dropped to AUC=0.537 in testing. For this reason, only properly validated results should be taken as indicators of actual performance. Thus only the testing results reported in this study should be taken as indicative of true performance.

The ASM was built by using the information along the boundary of target objects to represent the variation of shape only, while the AAM takes advantage of shape as well as the intensity information of the region within a target object to represent variation in appearance. A number of principal modes in the shape and appearance models were created respectively using PCA to represent the variation of shape and appearance in training sets. Our finding showed that 95 percent of shape variations could be explained by the first six principal modes of shape. In contrast, the first six principal modes of appearance described a mere 27% of the variation in appearance. The first 50 principal modes of appearance were needed to achieve the same level of variation description in appearance. We compared the classification performance between the ASM and AAM using the first six principal modes of variation, and found that AAM produced better discrimination (AUC =0.707) than the ASM (AUC =0.658). In addition, we compared the best combination of six principal modes selected by exhaustive search for the ASM and AAM based on AUCscore, and found that the features suitable for representation of target subjects may not always be the features selected by exhaustive search for classification between the two groups. This illustrates the fact that, while PCA finds features (called principal modes) that are optimal for representing information, these features are not necessarily optimal for classification (Hastie et al. 2009).

In AAM the spacing within the region of the femur is about 6.5 smaller than the spacing along the boundary. This is because the bone has a smooth shape at the scale of a few millimetres (more accurately, the variation of shape at a smaller scale was not expected to contribute to fracture risk), but within the interior of the region, texture due to bone structure and/or ligaments will likely vary at a smaller scale.

Two texture analysis techniques, Gabor filters and textons, were chosen in this study to extract a set of texture features on several ROI. The discriminating accuracy using the features derived from the whole femoral neck region was found to be higher than using other regions. Furthermore, we found that the features from Gabor filters on the whole femoral neck region gave the best prediction of factures among all texture methods. Gabor filters are a powerful tool for extracting spatial orientation so that patterns of oriented structure may be recognised even if they appear indistinct to the human eye. This might explain the reason why the discriminating abilities using features extracted by Gabor filters on the whole neck region outperformed other regions in the proximal femur.

Implementing Gabor filters requires choosing an array of parameters. The choices made here were based on the literature and our own understanding of image characteristics of the particular data set, including pixel size and anticipated scope of texture. Further investigation regarding the effect of filter parameters on risk assessment was not undertaken for two reasons. First, previous experience with Gabor filters indicates that small variations in filter parameters do not effect outcomes in medical image analysis problems substantially. Second, the aim of the study was to ascertain if image texture analysis is able to improve risk assessment based on DXA image. This has been accomplished. The size and quality of the data set precluded accurate reporting of the full potential of texture analysis. Thus future work will include more in-depth texture analysis, well beyond minor refinements of Gabor filters, applied to full resolution DXA images from a diverse population.

#### Textural analysis of bone in assessment of fracture risk is a topical area; trabecular bone score is a recently developed tool that performs grey-level texture analysis on lumbar spine DXA images providing information relating to trabecular micro-architecture. Trabecular bone score has been shown to relate to fracture risk independent of clinical risk factors and aBMD, and has predictive value for fracture independent of fracture probabilities using FRAX (Harvey et al. 2015). The methodology discussed in this study provides information at the other site assessed by DXA, the femoral hip.

In conclusion we have identified and assessed methodology that might be applied to derive further information from routinely collected DXA scan data. In a modest sample, texture features derived from Gabor filters in combination with total T-score provided better estimates of risk than the standard measures of aBMD or total T-score alone. Estimates of risk were more accurate when shape and texture were measured on the whole femoral neck than other regions.

**Acknowledgments**

We would like to thank Rebecca Ross and Ahmad Elmansouri for their assistance with data transfer for this study. We would also like to thank the participants of the Hertfordshire Cohort Study and their General Practitioners. The Hertfordshire Cohort Study is funded by the Medical Research Council and Arthritis Research UK.

**References**

1. [Alonso CG](http://www.ncbi.nlm.nih.gov/pubmed/?term=Alonso%20CG%5BAuthor%5D&cauthor=true&cauthor_uid=11095176), [Curiel MD](http://www.ncbi.nlm.nih.gov/pubmed/?term=Curiel%20MD%5BAuthor%5D&cauthor=true&cauthor_uid=11095176), [Carranza FH](http://www.ncbi.nlm.nih.gov/pubmed/?term=Carranza%20FH%5BAuthor%5D&cauthor=true&cauthor_uid=11095176), [Cano RP](http://www.ncbi.nlm.nih.gov/pubmed/?term=Cano%20RP%5BAuthor%5D&cauthor=true&cauthor_uid=11095176), [Peréz AD](http://www.ncbi.nlm.nih.gov/pubmed/?term=Per%C3%A9z%20AD%5BAuthor%5D&cauthor=true&cauthor_uid=11095176). 2000. Femoral bone mineral density, neck-shaft angle and mean femoral neck width as predictors of hip fracture in men and women. Osteoporosis Int. 11:714-720.
2. Baker-Lepain JC, Luker KR, Lynch JA, Parimi N, Nevitt MC, Lane NE. 2011. Active shape modeling of the hip in the prediction of incident hip fracture. J Bone Miner Res. 26:468-474.
3. [Beck TJ](http://www.ncbi.nlm.nih.gov/pubmed/?term=Beck%20TJ%5BAuthor%5D&cauthor=true&cauthor_uid=11127194), [Looker AC](http://www.ncbi.nlm.nih.gov/pubmed/?term=Looker%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=11127194), [Ruff CB](http://www.ncbi.nlm.nih.gov/pubmed/?term=Ruff%20CB%5BAuthor%5D&cauthor=true&cauthor_uid=11127194), [Sievanen H](http://www.ncbi.nlm.nih.gov/pubmed/?term=Sievanen%20H%5BAuthor%5D&cauthor=true&cauthor_uid=11127194), [Wahner HW](http://www.ncbi.nlm.nih.gov/pubmed/?term=Wahner%20HW%5BAuthor%5D&cauthor=true&cauthor_uid=11127194). 2000. Structural trends in the aging femoral neck and proximal shaft: analysis of the third national health and nutrition examination survey Dual-Energy X-ray absorptiometry data. J Bone Miner Res. 15:2297-2304.
4. Bela J. 1981. Textons, the elements of texture perception, and their interactions. Nature. 290:91-97.
5. Bredbenner TL, Mason RL, Havill LM, Orwoll ES, Nicolella DP. 2014. Fracture risk predictions based on statistical shape and density modeling of the proximal femur. J Bone Miner Res. 29:2090-2100.
6. Cooper C, Melton LJ. 1996 Magnitude and impact of osteoporosis and fractures. Osteoporosis. Academic Press, San Diego.
7. Cootes TF, Edwards GJ, Taylor CJ. 2001. Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence. 23:681-685.
8. Cootes TF, Hill A, Taylor CJ, Haslam J. 1994. Use of active shape models for locating structures in medical images. Image and Vision Computing. 12:355-365.
9. Cootes TF, Taylor CJ, Cooper DH, Graham J. 1995. Active shape models: their training and application. Computer Vision and Image Understanding. 61:38–59.
10. Cootes TF, Taylor CJ. 2001. Statistical models of appearance for medical image analysis and computer vision. Proc. SPIE Medical Imaging. 236-248.
11. Cummings SR, Black DM, Nevitt MC, Browner W, Cauley J, Ensrud K, [Genant HK](http://www.ncbi.nlm.nih.gov/pubmed/?term=Genant%20HK%5BAuthor%5D&cauthor=true&cauthor_uid=8093403), [Palermo L](http://www.ncbi.nlm.nih.gov/pubmed/?term=Palermo%20L%5BAuthor%5D&cauthor=true&cauthor_uid=8093403), [Scott J](http://www.ncbi.nlm.nih.gov/pubmed/?term=Scott%20J%5BAuthor%5D&cauthor=true&cauthor_uid=8093403), [Vogt TM](http://www.ncbi.nlm.nih.gov/pubmed/?term=Vogt%20TM%5BAuthor%5D&cauthor=true&cauthor_uid=8093403). 1993. Bone density at various sites for prediction of hip fractures. Lancet. 341:72-75.
12. Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, [Cauley J](http://www.ncbi.nlm.nih.gov/pubmed/?term=Cauley%20J%5BAuthor%5D&cauthor=true&cauthor_uid=7862179), [Black D](http://www.ncbi.nlm.nih.gov/pubmed/?term=Black%20D%5BAuthor%5D&cauthor=true&cauthor_uid=7862179), [Vogt TM](http://www.ncbi.nlm.nih.gov/pubmed/?term=Vogt%20TM%5BAuthor%5D&cauthor=true&cauthor_uid=7862179). 1995. Risk factors for hip fracture in white women. N Engl J Med. 332:767–773.
13. Dinçel VE, [Şengelen M,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+%c5%9eengelen%2c+Meltem+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe) [Sepici V,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+Sepici%2c+Vesile+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe" \t "_parent" \o "Find all records containing) [Çavuşoğlu T,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+%c3%87avu%c5%9fo%c4%9flu%2c+Turgay+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe" \t "_parent" \o "Find all records containing) [Sepici B](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+Sepici%2c+Beh%c3%a7et&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe" \t "_parent" \o "Find all records containing). 2008. The association of proximal femur geometry with hip fracture risk. Clinical Anatomy. 21:575-580.
14. Duboeuf F, Braillon P, Chapuy MC, Haond P, Hardouin C, Meary MF, Delmas PD, Meunier PJ. 1991. Bone mineral density of the hip measured with dual-energy X-ray absorptiometry in normal elderly women and in patients with hip fracture*.* Osteoporosis International. 1:242-249.
15. Faulkner KG, McClung M, Cummings SR. 1994. Automated evaluation of hip axis length for predicting hip fracture. J Bone Miner Res. 9:1065-1070.
16. Fogel I, Sagi D. 1989. Gabor filters as texture discriminator*.* Biological Cybernetics. 61:103-113.
17. Gangeh MJ, Sørensen L, Shaker SB, Kamel MS, de Bruijne M, Loog M. 2010. A texton-based approach for the classification of lung parenchyma in CT images. [Medical Image Computing and Computer-Assisted Intervention](http://link.springer.com/book/10.1007/978-3-642-15711-0). 6363:596-603.
18. Ginneken BV, Frangi AF, Staal JJ, Ter Haar Romeny BM, Viergever MA. 2002. Active shape model segmentation with optimal features. IEEE Transactions on Medical Imaging. 21:924-933.
19. Gnudi S, Ripamonti C, Gualtieri G, Malavolta N. 1999. Geometry of proximal femur in the prediction of hip fracture in osteoporotic women. Br J Radiol. 72:729-733.
20. Goodyear SR, Barr RJ, McCloskey E, Alesci S, Aspden RM, Reid DM, [Gregory JS](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gregory%20JS%5BAuthor%5D&cauthor=true&cauthor_uid=23220597). 2013. Can we improve the prediction of hip fracture by assessing bone structure using shape and appearance modelling? Bone. 53:188-193.
21. Gregory JS, Aspden RM. 2008. Femoral geometry as a risk factor for osteoporotic hip fracture in men and women. Med Eng Phys. 30:1275-1286.
22. Gregory JS, Stewart A, Undrill PE, Reid DM, Aspden, RM. 2005. Bone shape, structure, and density as determinants of osteoporotic hip fracture- A pilot study investigating the combination of risk factors. Investigative Radiology. 40:591-597.
23. Gregory JS, Testi D, Stewart A, Undrill PE, Reid DM, Aspden RM. 2004. A method for assessment of the shape of the proximal femur and its relationship to osteoporotic hip fracture. Osteoporosis Int. 15:5-11.
24. Gregory JS, Waarsing JH, Day J, Pols HA, Reijman M, Weinans H, Aspden RM. 2007. Early identification of radiographic osteoarthritis of the hip using an active shape model to quantify changes in bone morphometric features: Can hip shape tell us anything about the progression of osteoarthritis? Arthritis Rheum. 56:3634-3643.
25. Grigorescu, SE, Petkov N, Kruizinga P. 2002. Comparison of texture features based on Gabor filters*.* Image Processing, IEEE Transactions. 11:1160-1167.
26. Harvey NC, Glüer CC, Binkley N, McCloskey EV, Brandi M-L, Cooper C, Kendler D, Lamy O, Laslop A, Camargos BM, Reginster J-Y, Rizzoli R, Kanis JA. 2015. Trabecular bone score (TBS) as a new complementary approach for osteoporosis evaluation in clinical practice. Bone. 78:216-224.
27. Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning: data mining, inference and prediction. 2nd ed. New York: Springer.
28. Heaney RP, Barger-Lux MJ, Davies KM, Ryan RA, Johnson ML, Gong G. 1997. Bone dimensional change with age: interactions of genetic, hormonal, and body size variables. Osteoporosis Int. 7:426-431.
29. Kanis JA. 2002. Diagnosis of osteoporosis and assessment of fracture risk. Lancet. 359:1929-1936.
30. Kim, D, Sung J. 2009. Automated Face Analysis: Emerging Technologies and Research: Emerging Technologies and Research. IGI Global.
31. Kruizinga P, Petkov 1999. Nonlinear operator for oriented texture. IEEE transactions on image processing. 8:1395-1407.
32. Lespessailles E, Chappard C, Bonnet N, Benhamou CL. 2006. Imaging techniques for evaluating bone microarchitecture. Joint Bone Spine. 73:254-261.
33. Lim SE, Xing Y, Chen Y, Leow WK, Howe TS, Png MA. 2004. Detection of Femur and Radius Fractures in X-Ray Images. Proceedings of 2nd International Conference on Advances in Medical Signal and Information Processing; Singapore.
34. Lynch JA, Parimi, N, Chaganti RK, Nevitt MC, Lane NE. 2009. The association of proximal femoral shape and incident radiographic hip OA in elderly women. Osteoarthritis and Cartilage. 17:1313-1318.
35. Malik J, Belongie S, Leung T, Shi J. 2001. Contour and texture analysis for image segmentation. International Journal of Computer Vision. 43:7-27.
36. Marco DM, Fulvia V, Roberto DM, Fulvio M, Alberto C, 2004. Total lymphocyte count and femoral bone mineral density in postmenopausal women*.* Journal of bone and mineral metabolism. 22:58-63.
37. Marshall D, Johnell O, Wedel H. 1996. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. BMJ. 312:1254-1259.
38. Metz CE. 1978. Basic principles of ROC analysis. Seminars in Nuclear Medicine. 8:283-298.
39. Michelotti J, Clark J. 1999. Femoral neck length and hip fracture risk. J Bone Miner Res. 14:1714-1720.
40. [Pande I,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=Pande%2c+I.+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe) [O”Neill TW,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+O%e2%80%9dNeill%2c+T.+W.+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe" \t "_parent" \o "Find all records containing) [Pritchard C,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+Pritchard%2c+C.+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe) [Scott DL,](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+Scott%2c+D.+L.+&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe)[Woolf AD](http://flinders-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vl(freeText0)=+Woolf%2c+A.+D.&vl(15290103UI0)=creator&vl(31603427UI1)=all_items&fn=search&tab=default_tab&mode=Basic&vid=FUL&scp.scps=scope%3a(61FUL)%2cscope%3a(FUL_METALIB)%2cscope%3a(FUL_DSPACE)%2cprimo_central_multiple_fe). 2000. Bone mineral density, hip axis length and risk of hip fracture in men: results from the Cornwall hip fracture study. Osteoporosis Int. 11:866-870.
41. Petkov N. 1995. Biologically motivated computationally intensive approaches to image pattern recognition. Future Generation Computer System. 11:451-465.
42. Petroudi S, Kadir T, Brady M. 2003. Automatic classification of mammographic parenchymal patterns: a statistical approach. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Oxford, UK.
43. Pramudito JT, Soegijoko S, Mengko TR, Muchtadi FI, Wachjudi RG. 2007. Trabecular pattern analysis of proximal femur radiographs for osteoporosis detection. Journal of Biomedical and Pharmaceutical Engineering. 1:45-51.
44. Pulkkinen P, Partanen J, Jalovaara P, Jämsä. T. 2010. BMD T-score discriminates trochanteric fractures from unfractured controls, whereas geometry discriminates cervical fracture cases from unfractured controls of similar BMD. Osteoporosis Int. 21:1269-1276.
45. Robbins JA, Schott AM, Garnero P, Delmas PD, Hans D, Meunier PJ. 2005. Risk factors for hip fracture in women with high BMD: EPIDOS study. Osteoporosis Int. 16:149-154.
46. Sarkalkan N, Weinans H, Zadpoor AA. 2014. [Statistical shape and appearance models of bones](http://dx.doi.org/10.1016/j.bone.2013.12.006). Bone. 60: 129-140.
47. Shotton J, Johnson M, Cipolla R. 2008. Semantic texton forests for image categorization and segmentation. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition; University of Cambridge, UK.
48. Smyth PP, Taylor CJ, Adams JE. 1997. Automatic measurement of vertebral shape using active shape models. Image and Vision Computing. 15: 575-581.
49. [Stone KL, Seeley DG, Lui LY](http://www.uptodate.com/contents/osteoporotic-fracture-risk-assessment/abstract/14), Cauley JA, Ensrud K, Browner WS, [Nevitt MC](http://www.ncbi.nlm.nih.gov/pubmed/?term=Nevitt%20MC%5BAuthor%5D&cauthor=true&cauthor_uid=14606506), [Cummings SR](http://www.ncbi.nlm.nih.gov/pubmed/?term=Cummings%20SR%5BAuthor%5D&cauthor=true&cauthor_uid=14606506). 2003. BMD at multiple sites and risk of fracture of multiple types: long-term results from the study of osteoporotic fractures. J Bone Miner Res. 18:1947-1954.
50. Syddall HE, Sayer AA, Dennison EM, Martin HJ, Barker DJP, Cooper C. 2005. Cohort profile: The Hertfordshire Cohort Study. International Journal of Epidemiology. 34: 1234-1242.
51. Turner MR. 1986. Texture discrimination by Gabor functions*.* Biological Cybernetics 55:71-82.
52. Varma M, Zisserman A. 2005. A Statistical Approach to Texture Classification from Single Images. International Journal of Computer Vision. 62:61-81.
53. Vijay A, Shankar N, Ligesh C, Anburajan M. Evaluation of osteoporosis using CT image of proximal femur compared with dual energy X-ray absorptiometry (DXA) as the standard. Proceedings of the 2011 3rd International Conference on Electronics Computer Technology (ICECT), Kanyakumari, India.
54. Wilkin T, Devendra D, Dequeker J, Luyten FP. 2001. Bone densitometry is not a good predictor of hip fracture. BMJ. 323:795-797.
55. Yap DWH, Chen Y, Leow WK, Howe TS, Png MA. 2004. Detecting femur fractures by texture analysis of trabeculae. Proceedings of the 17th International Conference on Pattern Recognition; Univ. of Singapore, Singapore.
56. Yujia L, William DL, Yunhua L. 2015 Study of DXA-derived lateral–medial cortical bone thickness in assessing hip fracture risk*.* Bone Reports. 2:44-51.

**Table 1**

Baseline characteristics of total subjects (29 fracture cases and 90 control cases). Values shown as mean±SD, and *p* values are from t-test.

|  |  |  |  |
| --- | --- | --- | --- |
| Methods | Fracture group (n=29) | Control group (n=90) | *p* Values |
| Age (years) | 65.90±2.91 | 65.64±3.09 | 0.700 |
| Neck T-score | -1.17±0.81 | -0.78±1.33 | 0.140 |
| Total T-score | -0.70±0.87 | -0.08±1.10 | 0.0065\*\* |
| Total aBMD(g/cm2) | 0.88±0.12 | 0.96±0.15 | 0.0138\* |
| Neck aBMD(g/cm2) | 0.74±0.10 | 0.78±0.21 | 0.320 |
| HAL(mm) | 109.10±9.72 | 110.76±11.01 | 0.470 |
| NSA(⁰) | 128.93±5.18 | 128.58±5.86 | 0.770 |

\**p*<0.05, \*\**P*<0.01; HAL= hip axis length; NSA= neck shaft angle.

**Table 2a**

Performance of each technique in fracture discrimination in training and testing sets.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Best 1 feature | Best 3 features\* | Best 6 features\* |
| Total aBMD | Training | 0.651 | - | - |
| Testing | 0.699 | - | - |
| Neck aBMD | Training | 0.629 | - | - |
| Testing | 0.651 | - | - |
| Total Tscore | Training | 0.688 | - | - |
| Testing | 0.692 | - | - |
| Neck T-score | Training | 0.627 | - | - |
| Testing | 0.654 | - | - |
| ASM | Training | 0.610 | 0.715 | 0.778 |
| Testing | 0.402 | 0.475 | 0.563 |
| AAM | Training | 0.660 | 0.830 | 0.900 |
| Testing | 0.479 | 0.417 | 0.487 |

**\*** Where available – for ASM and AAM only

**Table 2b**

Performance of Gabor filters and Textons method with regard to fracture discrimination in training and testing sets (results reported as AUC)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gabor filters  | Neighbour textons  | Gabor textons  |
| Number of features selected | 1 | 3 | 6 | 1 | 3 | 6 | 1 | 3 | 6 |
| Whole hip | Training | 0.646 | 0765 | 0.841 | 0.698 | 0.794 | 0.873 | 0.626 | 0.760 | 0.821 |
| Testing | 0.394 | 0.426 | 0.429 | 0.429 | 0.433 | 0.398 | 0.501 | 0.394 | 0.455 |
| Whole femoral Neck | Training | 0.736 | 0.770 | 0.871 | 0.624 | 0.743 | 0.859 | 0.719 | 0.777 | 0.808 |
| Testing | 0.700\* | 0.588 | 0.603 | 0.504 | 0.413 | 0.388 | 0.674 | 0.478 | 0.515 |
| Neck | Training | 0.648 | 0.721 | 0.820 | 0.607 | 0.714 | 0.872 | 0.601 | 0.727 | 0.785 |
| Testing | 0.358 | 0.401 | 0.436 | 0.402 | 0.498 | 0.324 | 0.387 | 0.437 | 0.530 |
| Ward’s Triangle | Training | 0.621 | 0.726 | 0.807 | 0.607 | 0.711 | 0.804 | 0.616 | 0.719 | 0.810 |
| Testing | 0.352 | 0.381 | 0.439 | 0.563 | 0.481 | 0.420 | 0.406 | 0.462 | 0.468 |
| Narrow Neck | Training | 0.647 | 0.719 | 0.792 | 0.562 | 0.708 | 0.801 | 0.599 | 0.699 | 0.756 |
| Testing | 0.368 | 0.463 | 0.391 | 0.455 | 0.280 | 0.297 | 0.391 | 0.409 | 0.386 |
| Inter-trochanter | Training | 0.619 | 0.819 | 0.858 | 0.678 | 0.780 | 0.838 | 0.587 | 0.675 | 0.731 |
| Testing | 0.595 | 0.545 | 0.521 | 0.419 | 0.513 | 0.417 | 0.466 | 0.433 | 0.591 |
| Femoral shaft | Training | 0.596 | 0.690 | 0.745 | 0.618 | 0.741 | 0.807 | 0.591 | 0.659 | 0.707 |
| Testing | 0.437 | 0.464 | 0.504 | 0.558 | 0.424 | 0.422 | 0.480 | 0.629 | 0.537 |

 \*The best testing AUCscore.

**Table 3**

Performance of the combinations of the method proposed (ASM, AAM, Gabor or textons) and the standard method (Total aBMD or T-score) with regard to fracture discrimination in training and testing sets (results reported as AUC)

|  |  |  |
| --- | --- | --- |
|  | Total aBMD  | Total T-score  |
| Number of features selected | 1 | 3 | 6 | 1 | 3 | 6 |
| ASM  | Training | **0.651\*\*** | 0.736 | 0.787 | **0.688** | 0.770 | 0.809 |
| Testing | **0.699** | 0.591  | 0.592 | **0.692** | 0.633 | 0.598 |
| AAM  | Training | 0.660 | 0.819 | 0.901 | **0.688** | 0.841 | 0.907 |
| Testing | 0.480 | 0.518 | 0.524 | **0.692** | 0.558 | 0.492 |
| Gabor filters on whole hip  | Training | **0.651** | 0.745 | 0.847 | **0.688** | 0.752 | 0.844 |
| Testing | **0.699** | 0.483 | 0.524 | **0.692** | 0.518 | 0.517 |
| Gabor filters on whole femoral neck  | Training | 0.736 | 0.729 | 0.862 | 0.736 | 0.751 | 0.870 |
| Testing | 0.700 | 0.507 | 0.687 |  0.700 | 0.575 | 0.787\* |
| Neighbour textons on whole hip  | Training | 0.698 | 0.801 | 0.880 | 0.698 | 0.808 | 0.884 |
| Testing | 0.429 | 0.593 |  0.506 | 0.429 | 0.570 | 0.571 |
| Neighbour textons on whole femoral neck  | Training | **0.651** | 0.759 | 0.864 | **0.688** | 0.797 | 0.870 |
| Testing | **0.699** | 0.675 | 0.606 | **0.692** | 0.646 | 0.628 |
| Gabor textons on whole hip  | Training | **0.651** | 0.740 | 0.838 | **0.688** | 0.756 | 0.847 |
| Testing | **0.699** | 0.522 | 0.537 | **0.692** | 0.717 | 0.527 |
| Gabor textons on whole femoral neck  | Training | 0.719 | 0.783 | 0.816 | 0.719 | 0.796 | 0.827 |
| Testing | 0.674 | 0.514 | 0.583 | 0.674 | 0.513 | 0.541 |

\*The best testing AUCscore; \*\* Bold indicates that the best feature was either total aBMD or the total T-score (applies to single features only).

**Table 4**

Performance of the combinations of ASM (6 features), AAM (6 features), Gabor filters on whole femoral neck (6 features), total BMD and total T-Score (2 features) with regard to fracture discrimination in training and testing sets (results reported as AUC)

|  |  |
| --- | --- |
| Number of features selected | Total BMD and Total T-Score  |
|  1 feature |  6 features | 11 features |
| ASM, AAM and Gabor filters on whole femoral neck | Training 0.732 |  0.916 | 1.000 |
| Testing 0.705\* |  0.603 | 0.537 |

\*The best testing AUCscore.

**Table 5a**

Comparison of study parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Paper Aa | Paper Bb | Paper Cc | Present paper |
| Database | 26 fracture24 non-fracture(all females) | 182 fracture364 non-fracture(all females) | 168 fracture231 non-fracture(all females) | 29 fracture90 non-fracture(both genders) |
| Fracture types | Hip fractures | Hip fractures | Hip fractures | All fractures |
| Point-point error analysis | Yes | No | No | No |
| landmarks | 29 pointshead included | 72 pointshead & pelvis | 60 pointshead included | 44 pointshead excluded |

aGregory et al. 2004; bGoodyear et al. 2013; cBaker-Lepain et al. 2011

**Table 5b**

Comparison of training AUC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Paper Aa | Paper Bb | Paper Cc | Present paper |
| ASM | 0.81 (4features) | 0.57 (1 feature) | 0.81 (10features) | 0.78 (6 features)  |
| AAM | n/a | 0.57 (1 feature) | n/a | 0.90 (6 features) |
| Ward aBMD | 0.95 (1 feature) | n/a | n/a | n/a |
| Total aBMD | 0.63(1 feature) | 0.62 (1 feature) | n/a | 0.65 (1 feature) |
| Neck aBMD | 0.79 (1 feature) | n/a | 0.68 | 0.63 (1 feature) |
| ASM & Ward aBMD | 0.96 (5features) | n/a | n/a | n/a |
| ASM & Neck aBMD | 0.89 (5 features) | n/a | 0.84 (11 features)  | n/a |
| ASM & Total aBMD | n/a | 0.79 (6 features)  | n/a | 0.79 (6 features)  |
| Multi-combination | n/a | 0.65 (3 features)ASM & AAM & Total aBMD  | n/a | 1.00 (11 features)Gabor & ASM & AAM & Total aBMD  |

aGregory et al. 2004; bGoodyear et al. 2013; cBaker-Lepain et al. 2011

**Table 5c**

Comparison of testing AUC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Paper Aa | Paper Bb | Paper Cc | Present paper |
|  | n/a | n/a | n/a | 0.79Gabor filters on whole femoral neck & total T-score (6 features) |

aGregory et al. 2004; bGoodyear et al. 2013; cBaker-Lepain et al. 2011

**Figure caption:**

Figure 1. Regions of Interest (ROI) considered for computing texture features using Gabor filters and textons.

Figure 2**.** An example of 44 manually selected contour points of the boundary of the proximal femur.

Figure 3**.** An example of normalized proximal femur image showing locations of sample points for characterising intensity patterns. The locations shown are indicative only of the regular sampling pattern used. The actual number of pixels sampled (82,830) is too large to display on the image.

Figure 4**.** Examples of Gabor filter outputs at various frequencies and orientations. (Left) The original image; (Right) 24 Gabor-filtered images with four different spatial frequencies (1/16, 1/18, 1/26 and 1/32 from top to bottom) and six different orientations (left to right).

Figure 5**.** An example of local 3x3 neighbourhoods-based textons with clustering K = 20. (Left) The original image; (middle) Texton; (Right) The distribution of textons.

Figure 6. ROC curves for T-score and five selected Gabor filter features from the femoral neck using an independent test set (solid line) and leave-one-out cross validation (dashed line). Independent test AUC = 0.787, leave-one-out AUC = 0.762.