Thesis

Statistical Appearance Models for Fast and Automated Estimation of Proximal Femur Fracture Risk Using Finite Element Models

Nazli Sarkalkan[1]

[1]*Department of Biomechanical Engineering - Tissue Biomechanics and Implants, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology (TU Delft)*

Exam committee:

Prof. Dr. H. H. Weinans, TU Delft
Dr. Ir. A. A. Zadpoor, TU Delft
Dr. Ir. P. Tiso, TU Delft

ABSTRACT

Use of DXA-measured aBMD is the common method to predict osteoporotic hip fractures in clinical settings. However, taking only the changes in aBMD into consideration is not enough to explain the whole variety of low energy fractures. It is deemed essential to develop alternative methods that also reflect the influence of other parameters (e.g. shape of the anatomical structure, load conditions), which are known to be associated with fracture. Development of subject specific FE models is a powerful instrument for investigating bone strength in vivo and, thus, for estimating the risk of fracture. As the mentioned alternative methods need to be adaptable to clinical settings while also being accurate and sufficiently fast for specific tasks, e.g. estimation of proximal femur fracture load, the main aim of this study was to develop a framework that is adaptable to clinical uses and has room for improvement. The presented semi-automatic framework covers development of patient specific FE models based on DXA to predict proximal femur fracture load. Information on the proximal femur shape of individuals were directly derived from DXA by Active Appearance Models (AAM), which detects the object of interest by fitting statistical shape models to the new set of images. To build up AAM, a training data set of DXA scans of 70 proximal femurs was used. Furthermore, 17 DXA scans of the proximal femurs that had been not included in the training set were used as test samples, on which the FE models were developed. To evaluate the effect of segmentation in prediction of proximal femur fracture load, two different cases were considered: proximal femurs that had been segmented using AAM and the same samples with manual segmentation. In order to evaluate the accuracy of AAM, leave-one-out experiments were conducted which provided a point-to-curve error of 1.2470 ± 0.6505 (mm) (with 95% confidence). On the other hand, point-to-curve error in segmentation of 17 proximal femurs that were used in the FE analyses was computed as 1.4169 ± 0.7499 (mm) (with 95% confidence). Taking all of the 17 proximal femur samples into account, the

fracture loads were estimated to be 3870.9 ± 932.83 (N) for manual segmentation case and 3804.2 ± 850.11 (N) for segmentation case using AAM. A strong correlation was observed between these estimated failure loads (R^2 =0.8197). On the other hand, it was noticed that even small errors (e.g. 1.06 mm) in segmentation process might result in larger errors (e.g. 24.1%) in the prediction of fracture load.

This work presents the first results obtained with the created framework, which is found to perform sufficiently well compared to its equivalents and is easily adaptable to clinical settings. However, considering the load prediction sensitivity to segmentation, further improvement in the accuracy of the segmentation process is believed to be a vital step for future studies. Such a development might be valuable for the prediction accuracy of proximal femur fracture risk.

KEY WORDS: Osteoporosis, Active Appearance Models, Finite Element Analysis, Proximal Femur, Proximal Femur Fracture Load

1. Introduction

Every year, more than 8.9 million osteoporotic fractures occur around the world with one third taking place in Europe and resulting in 2 million disability-adjusted years (DALYs) (Strom, Borgstrom et al. 2011). Osteoporotic fractures, whose prevalence increases with age, mostly occur at hip, spine and distal forearm (Roberts, Pacheco et al. 2010). Among these, hip fractures are the most serious as they necessitate hospitalization and surgical intervention. To lessen their occurrences, early identification of individuals, who are at high risk of developing fracture, is essential. Since by designation of individuals in advance, fracture prevention strategies (e. g. use of drug that increases bone density) can be followed (Naylor, McCloskey et al. 2013).

In order to diagnose osteoporosis and to predict fractures, several densitometric techniques (e.g. dual-energy x-ray absorptiometry - DXA, quantitative computed tomography - QCT) that enable evaluation of the quantity and quality of the bone, have been developed (Lochmuller, Zeller et al. 1998, Strom, Borgstrom et al. 2011). Among these techniques, DXA that measures the bone mass per unit area (areal bone mineral density, aBMD, g/cm^2) is the most widely used in clinical settings (Whitmarsh, Fritscher et al. 2012).

Although use of aBMD measured with DXA is adequate to diagnose osteoporosis, it is not possible to explain the whole variety of low energy (e. g. fall from standing height or less) fractures by changes in aBMD (Vaananen, Jurvelin et al. 2012). This is due to the fact that bone mass is not the only parameter which affects the bone strength; there are many other factors (e. g. shape of the anatomical structure, loading conditions) that should be considered in fracture assessments (Testi, Viceconti et al. 2002, Naylor, McCloskey et al. 2013). With the intent of improving the fracture risk estimation of the densitometric techniques, several attempts have been made to develop methods that can take other risk factors (e. g. shape of the anatomical structure) into consideration (Keyak, Rossi et al. 1998). One of the endeavors

to evaluate fracture risk is the development of subject-specific finite element (FE) models (Schileo, Taddei et al. 2008). These models are powerful tools to predict the bone strength in vivo, since they include information about the shape of the bone, bone tissue material properties and the external loading conditions. To develop subject-specific FE models, images that are acquired by various imaging modalities (e. g. CT, MRI for 3D models; DXA for 2D models) can directly be used (Poelert, Valstar et al. 2013). With the aim of generating FE models from images, objects of interests need to be segmented. Thus, segmentation process is one of the essential steps of subject-specific FE model generation. The segmentation procedure should be accurate enough to obtain realistic information on the shape of the interested structure, consequently to get more representative FE models. Besides, to decrease the time consumed on the preparation of FE models, segmentation process should be fast and should require less user interaction.

The aim of this study was to develop a fast and automated (i.e. semi-automatic) frame work to estimate the risk of proximal femur fracture. The process mainly covers the use of Active Appearance Model (AAM) (Cootes, Edwards et al. 1998) which is a wide known method for segmentation technique based on deformable models, and the 2D FE modeling of the segmented proximal femur. In the methods section, the different steps taken in the framework (*Figure 1*) are presented. Afterwards, in the results section, outputs of leave-out-experiments that were conducted for evaluation of AAM's performance in segmentation and of 2D FE analyses are given. Finally, in the discussion section, the results of the study will be evaluated and recommendations to further develop the framework will be given.

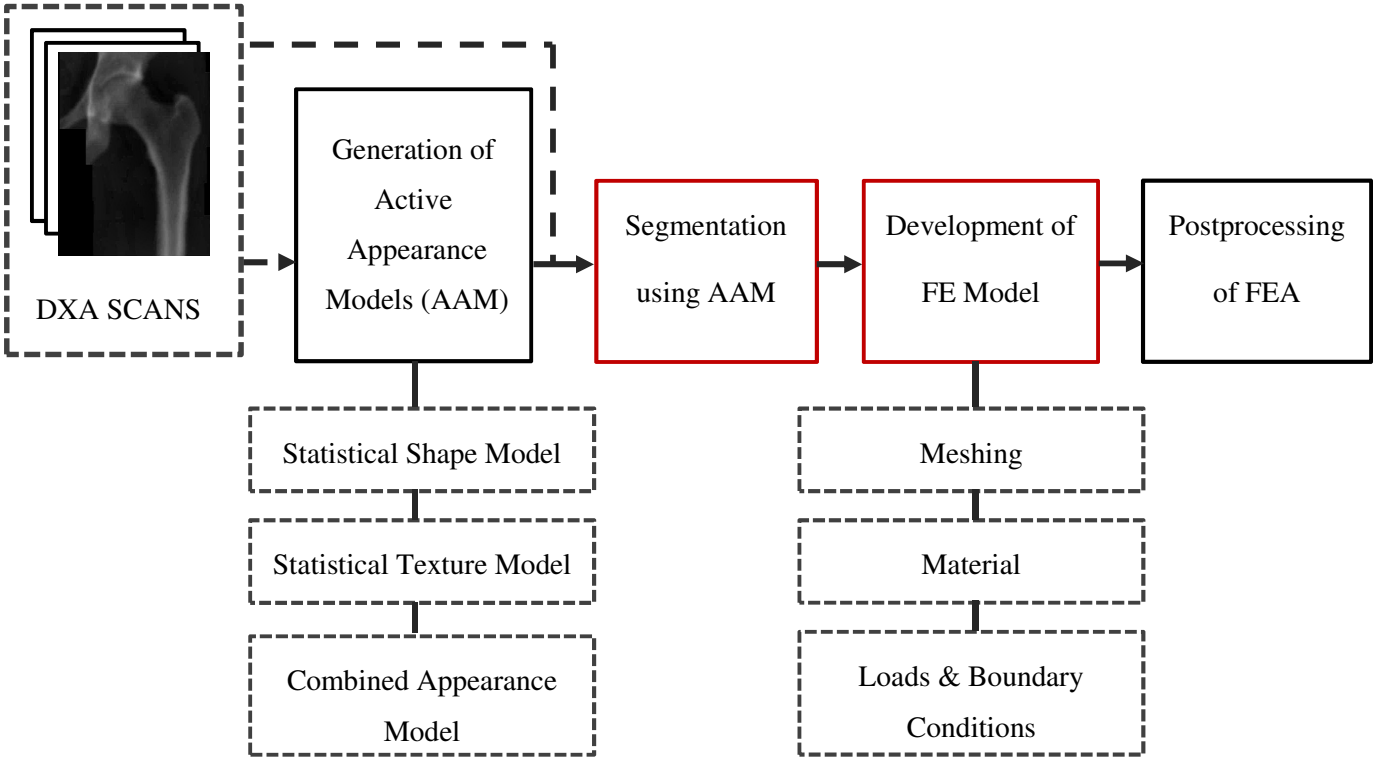

Figure 1: Main Framework to predict proximal femur fracture load.

2. Materials and Methods

The framework to predict proximal femur fracture load is composed of two main steps: segmentation of proximal femurs from DXA scans by using AAM and the development of subject–specific FE models. In order to search images and to take out the objects of interest (i.e. proximal femur), the generation of AAM is required. On the other hand, to evaluate the FE analyses, post-processing is necessary.

In the following subsections; the general information about the data (i.e. DXA scans of osteoporotic patients) used within the study, the steps involved in development of AAM (generation of statistical shape model, statistical texture model and of combined appearance model), in segmentation of proximal femur and in development of FE model (meshing, assignment of material, loads and boundary conditions) are presented. In addition, the methods to evaluate the performance of AAM in segmentation and the FE model to predict proximal femur fracture are explained.

Through all the steps, MatLab programming environment (The MathWorks Inc., USA) was used. Moreover, the FEA software Abaqus 6.13 (Dassault Systèmes Simulia Corp., USA) was used to generate mesh on segmented proximal femurs and employed as solver.

2.1. Data

Anonymous DXA scans (Lunar Prodigy, GE Healthcare - USA) of 151 osteoporotic patients were gathered from the Erasmus Medical Center (Rotterdam, The Netherlands). Some DXA scans were excluded when proximal femurs were not adequately imaged (e. g. scans including foreign objects such as buttons), and when the patients received implants. Following the elimination, DXA scans of 87 osteoporotic patients remained and a single hip image (left or right femur) for each subject was retrieved from these scans. Furthermore, 87 DXA scans were randomly divided into two groups; one including 70 DXA scans and one with 17 DXA scans. The first group is composed of the DXA scans of patients (56 Female and 14 Male) with an average age of 61.5 ± 12.6 years and this set was used in AAM generation. On the other hand, the second group that covers DXA scans of patients (12 Female and 5 Male) with an average of 61.82 ± 10.7 years was used in segmentation procedure, thus, in the development of FE models to predict hip fracture load.

2.2. Statistical Model

The generated statistical model was based on Active Appearance Model (Cootes, Edwards et al. 1998). This model basically captures the shape and texture (i.e. the pixel intensities across the object, thus the aBMD distribution) variations within a studied population. As the main aim of the statistical models is to reflect variations, the model needs to be trained with a number of examples of the object of interest (i.e. proximal femur) (Bryan, Nair et al. 2009). The basic steps involved in the development of AAM from 70 DXA scans of proximal femurs are: generation of statistical shape model, of statistical texture model and unification of these models into one complete and compact model: combined appearance model. These steps will

briefly be presented in the following subsections. On the other hand, detailed information on the main methods is given in Appendix I. For AAM generation, the current framework adopted the general approach, which can be found in MATLAB File Exchange, developed by Dirk-Jan Kroon.

2.1.1. Statistical Shape Model

The procedure for generation of statistical shape models is mainly composed of shape representation and statistical analysis of the shape data. With the intent of representing shapes of proximal femurs, the most generic method that requires definition of a set of points (landmarks) was used (Heimann and Meinzer 2009). Before going through the definition of a set of points, image contrast enhancement is performed to improve pictorial information of the DXA scans for better visual interpretation (*Figure 2*). To enhance the image contrast, the Min-Max filter method described by Westerweel was used (Westerweel 1993) (see Appendix I for further information).

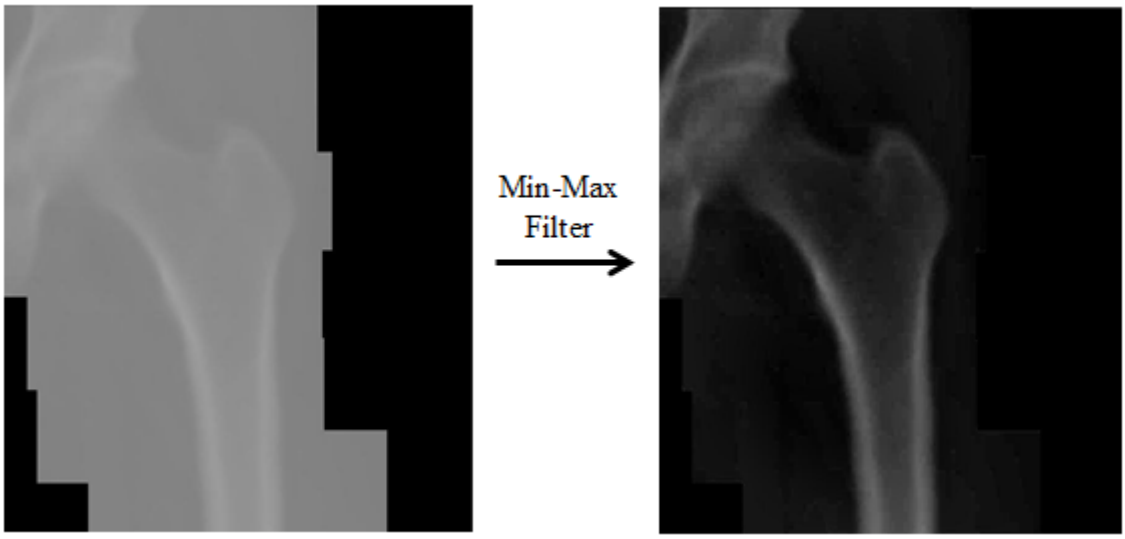

Figure 2: Application of Min-Max filter on an original DXA scan to enhance image contrast for better visual interpretation.

Following the contrast enhancement of images, a total number of 70 landmarks were manually placed along the contour of proximal femurs. As the lesser trochanter was not visible on numerous images, the lesser trochanter region was not considered (*Figure 3*).

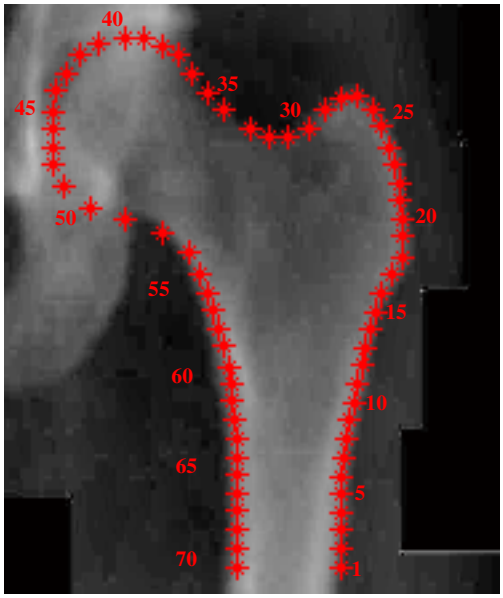

Figure 3: 70 landmarks were placed along the contour of proximal femur.

Subsequently, coordinates of all points were concatenated into a vector, **x** that represents the shape:

$$\mathbf{x} = \left(x_1, \ldots, x_n, y_1, \ldots, y_n\right)^T \tag{1}$$

The annotation procedure was repeated for each image of the training data, i.e. 70 DXA scans. During this process, care was taken to place all the points of a shape in a corresponding manner with those of the other remaining images of the training dataset.

In order to obtain true shape representation, all the shapes were aligned in a common coordinate system (*Figure 4*) by using the Generalized Procrustes Analysis (GPA) (Goodall 1991). Translational effects were removed by positioning all the shapes in such a way that their centroids were at the origin (Pegg, Mellon et al. 2012) . The scaling and the rotational effects were removed by using the method described by Du et al. and Arun et al., respectively (Arun, Huang et al. 1987, Du, Zheng et al. 2007) (see Appendix I).

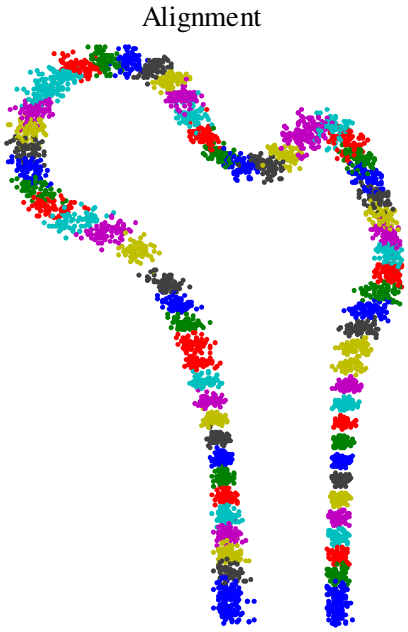

Figure 4: Alignment of 70 proximal femur shapes into a common coordinate frame.

After alignment of all the shapes into a common coordinate system, the intra-class shape variation was extracted by performing Principal Component Analysis (PCA) (Jolliffe 2005). The principal modes of variation (eigenvectors), Φ_s, and their respective variances (eigenvalues), λ_s, were found by performing singular value decomposition (SVD) on the aligned data (Heimann and Meinzer 2009). Subsequently, a shape instance can be described by adding contributions of the most important modes of variation to the mean shape:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi}_s \mathbf{b}_s \tag{2}$$

where, $\mathbf{b}_s$ is a set of shape parameters and $\mathbf{\Phi}_s$ represents the matrix of ordered eigenvectors. *Figure 5* shows the mean shape and the deformation of the mean shape using the first three modes of variation of the shape model.

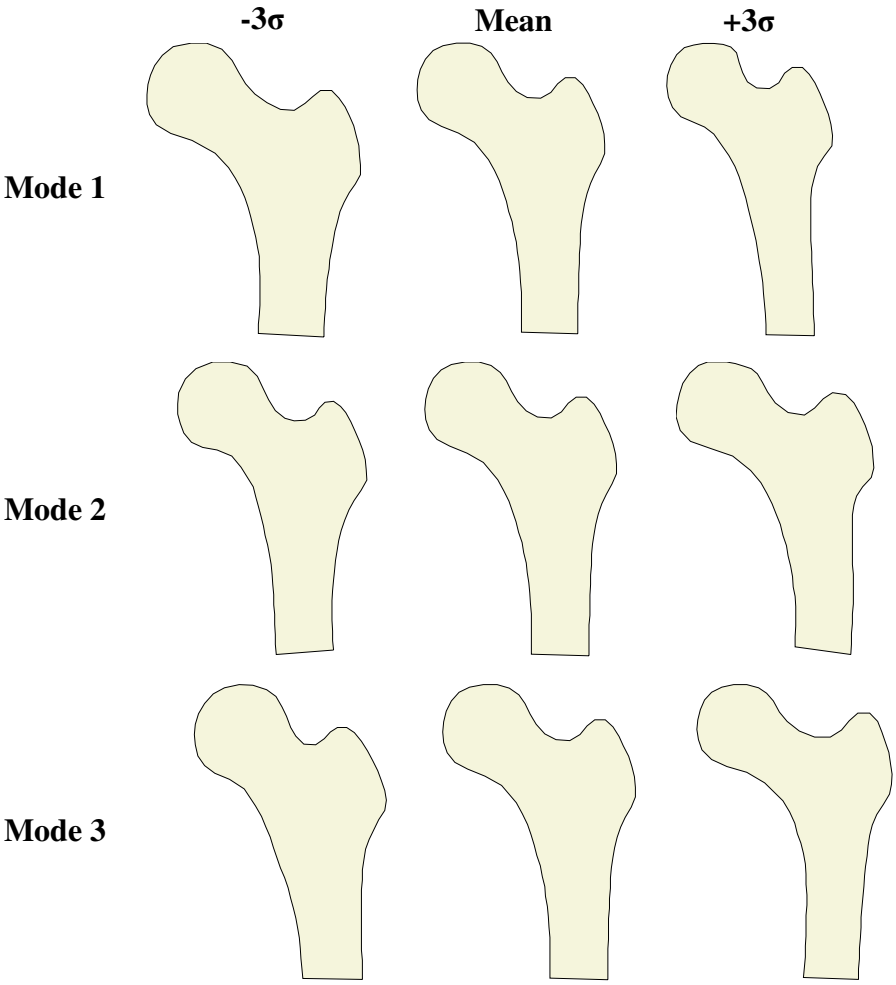

Figure 5: Mean shape and mean shape deformation using the first three modes of variation of the shape model.

To determine how many modes to be retained, the number of modes was increased until the ratio of the accumulated variance to the total variance, *r*, reaches 0.98 (Heimann and Meinzer 2009). 98% of the shape variation can be modeled with 16 modes (*Figure 6*). On the other hand, an important issue when generating a shape using Eq. 2 is to limit its variation to obtain plausible shapes. Therefore, vector $\mathbf{b}_s$, which contains the shape model parameters, has to be confined within a certain range (Heimann and Meinzer 2009). A common approach to generate shapes similar to those found in the original training dataset was considered by limiting the model parameters as following:

$$-3\sqrt{\lambda_i} \le b_i \le 3\sqrt{\lambda_i} \tag{3}$$

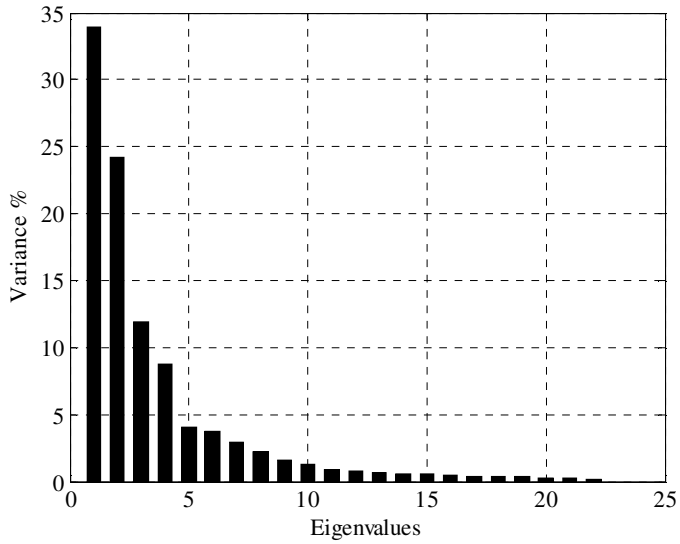

Figure 6: Shape eigenvalues in descending order.

2.1.1. Statistical Models of Texture

With the aim of building the statistical models of texture as described by Cootes et al. (Cootes, Edwards et al. 2001), there are three main steps to be followed. These steps include the capture of pixel information, the normalization of the data to discard the global illumination effects and the modeling of the normalized texture variation (Stegmann 2000). In order to capture texture information (i.e. pixel intensities across the proximal femur), a piece-wise affine warp based on the Delaunay triangulation was used (Stegmann 2000) (see Appendix I). Each training example was warped to the mean shape. In this way, the spurious texture variations that might rise due to shape differences were removed (Cootes and Taylor 2004). The intensity information from the shape normalized image was sampled using bilinear interpolation (see Appendix I) to form a texture vector, **g**. Afterwards, the texture vector was normalized according to Cootes et al. (Cootes, Edwards et al. 2001):

$$\mathbf{g}_{norm} = \frac{\left(\mathbf{g} - \mu_g \mathbf{1}\right)}{\sigma_g} \tag{4}$$

where, **1** is a unit vector, μ_g and σ_g^2 are respectively the mean and variance of elements of a texture vector, **g**. Following the normalization of texture data, as in the case of statistical

shape models, the texture variation was extracted by applying PCA to the texture data. In a similar manner to shape instance generation, a texture instance can be created by deforming the mean texture by a linear combination of eigenvectors (Stegmann 2000):

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{\Phi}_g \mathbf{b}_g \tag{5}$$

where $\mathbf{b}_g$ is a set of texture parameters and $\mathbf{\Phi}_g$ represents the matrix of ordered eigenvectors. Within the study, as the deformable model might favor to remain inside the proximal femur (less accurate fit of the bone shape), the method described by Stegmann was used to deal with this problem (Stegmann 2000) (see Appendix I). Referring to the method, an extra region that was in the neighborhood of the proximal femur was included.

Figure 7 shows the mean texture and the deformation of the mean texture using the first three modes of variation of the texture model. By using 50 modes, 98% of the texture could be explained (*Figure 8*).

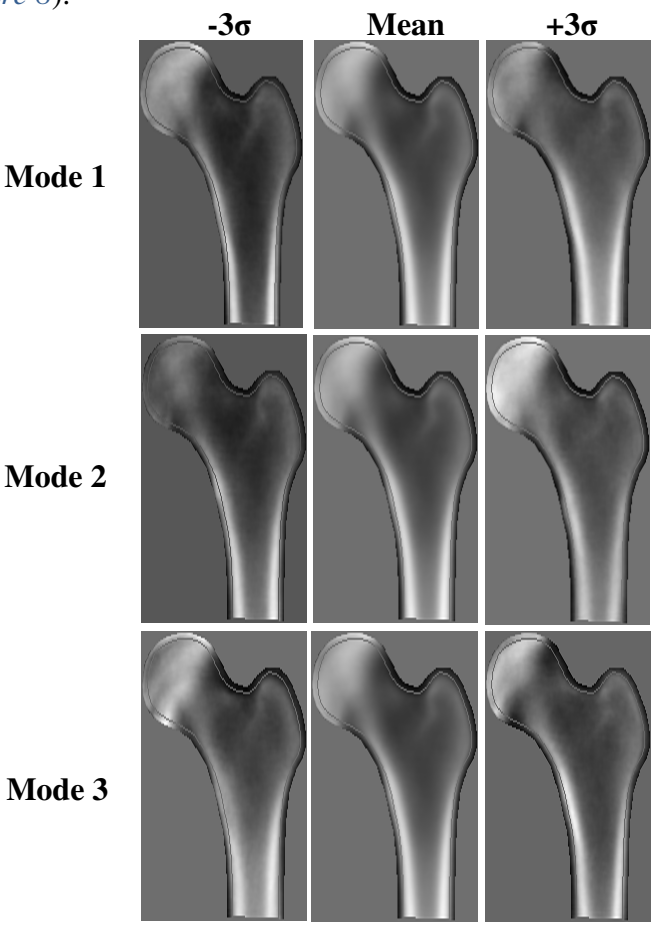

Figure 7: Mean texture and mean texture deformation using the first three modes of variation of the texture model.

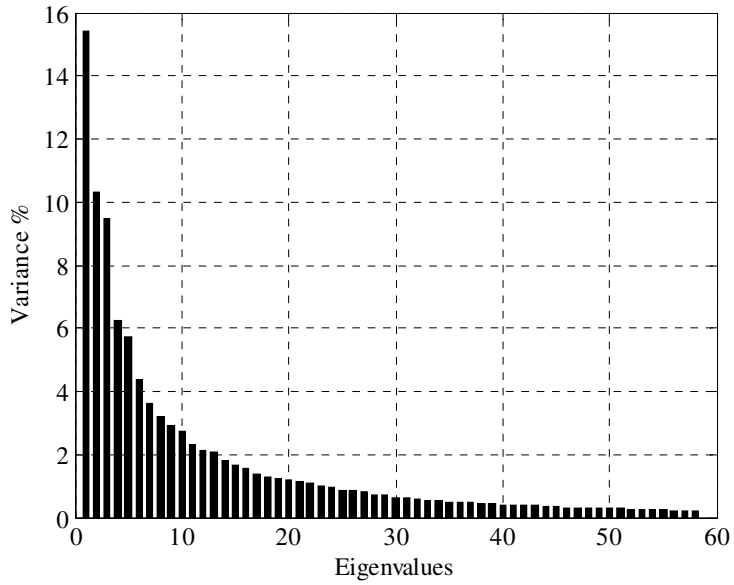

Figure 8: Texture eigenvalues in descending order.

2.1.2. Combined Appearance Model

With the aim of obtaining one complete compact model, the shape and texture models that were briefly presented, were unified (see Appendix I) (Cootes and Taylor 2004). Briefly, for each example of the training set, the shape and texture parameters were concatenated into a vector, **b**. The PCA, as in the case of statistical shape models, was applied on the data resulting in a further model:

$$\mathbf{b} = \mathbf{\Phi}_c \mathbf{c} \tag{6}$$

where $\mathbf{\Phi}_c$ represents the matrix of ordered eigenvectors and **c** is a set of parameters that control both the shape and texture of the model. As the combined model parameters are related to those of the shape model and of the texture model, these parameters can be found and can be substituted in Eq. 2 and 4 to generate new shape and texture instance (see Appendix I) (Whitmarsh, Humbert et al. 2011). On the other hand, 98% of the combined variation can be modeled with 44 parameters (*Figure 9*).

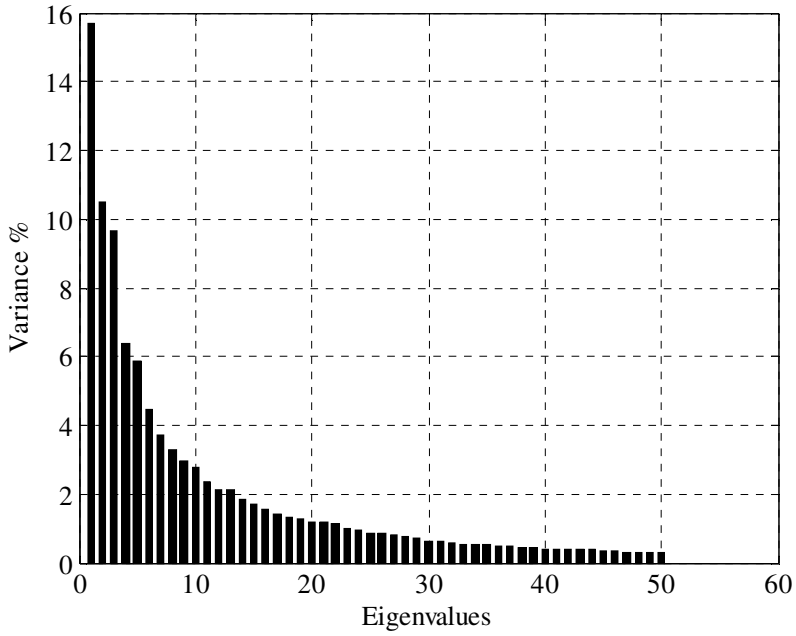

Figure 9: Combined appearance eigenvalues in descending order.

2.3. Search Algorithm

The main aim at building up AAM was to use it in DXA image search to detect the proximal femur. The search process can simply be considered as an optimization problem. In order to match the model to the image, the set of parameters that refer to the position and the shape should be found in such a way that the cost function, quality of the fit, is optimized (Cootes et al., 2004). AAM mainly takes the error between the model texture and image texture into account to evaluate the quality of fit (see Appendix I). During the search of images, the combined model parameters and pose parameters are estimated in such a way that the error is minimized. As there are quite a number of parameters (model and pose parameters) that should be adjusted during fitting, the optimization problem is high-dimensional. It is a known fact that solving high-dimensional problems are computationally expensive. In order to lessen the complexity of the problem, Cootes et al. had proposed the use of a-priori knowledge on properly adjusting the parameters during search (Cootes, Edwards et al. 2001). Within the study, the method described by Cootes et al., was used to lessen the complexity of problem. This method allows one to correct parameters during fitting. The details on getting a priori

knowledge and using it during matching of the model to the object is described in Appendix I. Briefly, the presented approach assumes that there is a linear relationship between the parameter updates (combined model and pose parameters), $\delta\mathbf{p}$, and the texture residuals, $\mathbf{r}(\mathbf{p})$, over the entire search (Heimann and Meinzer 2009, Kroon 2011):

$$\delta\mathbf{p} = -R\mathbf{r}(\mathbf{p}) \tag{7}$$

where, R is the derivative matrix. This was computed at the training time and used to correct parameters during matching. On the other hand, another important issue was the initialization of the search process. As the AAM search is one of the local search methods , it initially necessitates for a rough estimation of the correct solution (Heimann and Meinzer 2009). In this study, a rough estimation was supplied by positioning the mean shape close to the proximal femur.

2.4. Evaluation of AAM

In order to assess how well the developed AAM will perform, leave-one-out experiments were conducted on all training sets (70 DXA scans). With respect to these experiments, AAM were generated using all the training examples, but each time leaving one of the examples out. Following the generation of AAM, the model was fit to the missed out example. This procedure was repeated for all the training images. The segmentation error was found for each image by comparing the segmented shape to the corresponding known ground truth (i.e. same proximal femur that was manually segmented). The point-to-curve error (Stegmann 2000) was recorded for each image of the training set. On the other hand, in order to further evaluate the performance of AAM in segmentation, a point-to-curve error was computed considering the segmentation of 17 proximal femur samples that had not been included in the training set. These 17 proximal femur samples were used in development of finite element models. The point-to-curve error was found by comparing the proximal femurs segmented using AAM to their related known ground truth. Furthermore, to find out whether it is worthy to use AAM in

segmentation of proximal femurs, inter-observer reliability in manual segmentation was checked by asking three other users to put landmarks along the contours of 10 randomly selected proximal femurs. These were chosen from the set used in FE analyses. The proximal femur contours that were manually segmented by each user, were compared with their corresponding known ground truth (i.e. proximal femurs that had been manually segmented by the author).

2.5. Generation of Subject – Specific FE Models

With the aim of developing subject-specific FE models, information on the specific geometry, on the bone tissue material properties and on the loading conditions of individuals need to be supplied (Poelert, Valstar et al. 2013). With respect to specific geometry, the information was directly derived from the DXA scans by segmenting the proximal femur of each individual with the method described above. Afterwards, mesh was generated on the segmented bone. Proper material properties, loads and boundary conditions were assigned to the segmented bone. In the following subsections, the details of meshing procedure, definition of material properties, loads and boundary conditions and assignments of these to the segmented bone, are presented in detail.

2.5.1. Meshing

To generate mesh on the segmented bone, the extracted geometry information was sent to the FEA software Abaqus 6.13. Using the meshing tools of the software, bone model was discretized with the element type of CPS6M (6 node modified plane stress element). Before going through the FE analyses of 17 proximal femurs that were not involved in the training phase of AAM, mesh convergence test was performed by using one of the segmented bones. This test was necessary to be sure of obtaining mesh independent FE analyses results. Within the scope of mesh convergence study, two different cases were considered. In the first case, the spatial material distribution was not included. The same homogenous material parameters

were assigned to each element (E=6000MPa, ν=0.3) (Scholz, Hoffmann et al. 2013). In the second case, non-uniform distribution of material properties within the proximal femur was taken into consideration. To reach the goal, the material properties were defined with respect to gray values derived from the part of the original DXA image that had remained inside the segmented proximal femur (see Material assignment section). Defined material parameters were assigned to each element. During the mesh convergence study, mesh density was increased six times. The total number of elements and of nodes for each mesh refinement, are presented in *Table 1*.

Table 1: Total number of nodes and elements for each mesh refinements.

Mesh	Total Number of Nodes	Total Number of Elements
Mesh 1	1144	533
Mesh 2	4297	2070
Mesh 3	15215	7458
Mesh 4	18519	9094
Mesh 5	26008	12807
Mesh 6	57980	28695
Mesh 7	88229	43748

Five different elements were defined in coarsest mesh (*Figure 10*). The correspondent ones were determined for each mesh refinement.

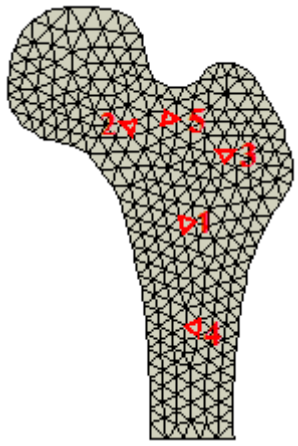

Figure 10: Five elements (red triangles) that were defined in the coarsest mesh.

The convergence was assessed based on the Von Mises stress (at element centroid) and the displacement (at nodes attached to the interested element) values. After evaluating the results of the mesh convergence study (see Results section), mesh density (i.e. Mesh 6) that would be used throughout the FE analyses of 17 proximal femurs was defined.

2.5.2. Material Assignments

The steps involved in material definition and assignment are described in detail, in Appendix II. Briefly, the centroid of each mesh element was found. For each element, a search region was defined. Using this region, the search area in the original DXA scan to find which pixels were staying within the mesh element, was narrowed. The gray values of pixels that remained inside the element were averaged. The mean gray value was used to calculate aBMD (*Figure 11*). Afterwards, aBMD was converted to volumetric BMD (vBMD) by using the subject-specific thickness of proximal femur (see Appendix II) and the apparent density in a manner described by Schileo et al. (Schileo, Taddei et al. 2008).

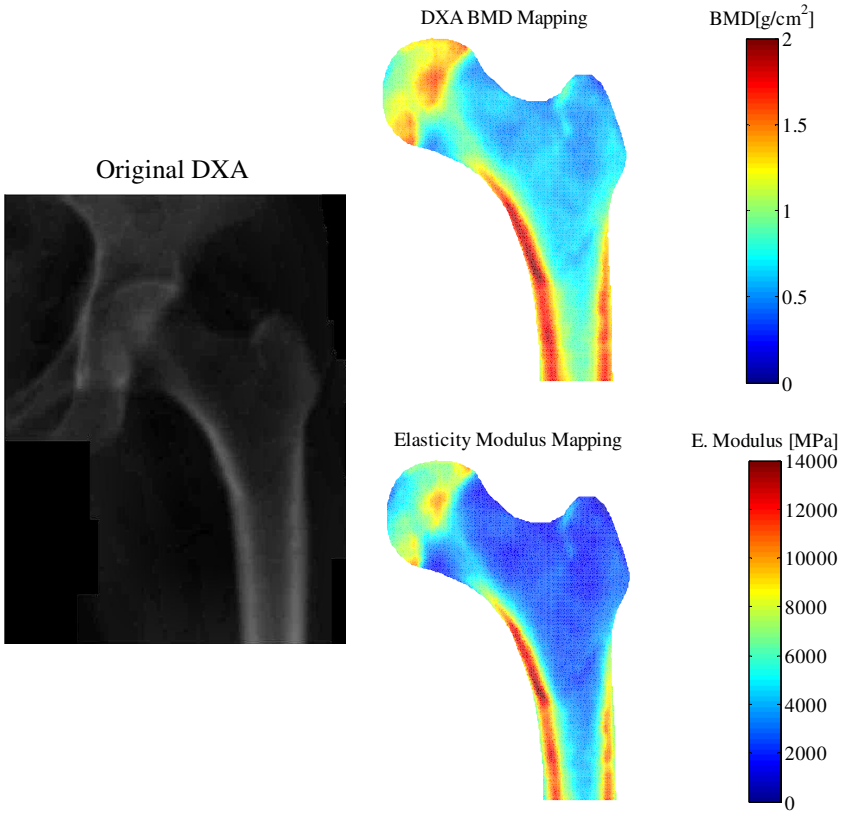

Figure 11: Original DXA scan (left) and definition of aBMD (g/cm^2) and Elasticity Modulus (MPa) using pixel intensities.

The material properties were defined by using empirical equations that relate the apparent density to elasticity modulus (*Figure 11*). The density-elasticity relationship (*Table 2*) that had been described by Morgan et al. (Morgan, Bayraktar et al. 2003) was preferred, since it was found to be best among the commonly used empirical equations to describe material properties in FE models (Scholz, Hoffmann et al. 2013).

Table 2: Equations used to define elasticity modulus, E (MPa), compressive strength, S_c (MPa) and tensile strength, S_t (MPa) from apparent density (g/cm^3).

Relationship	Density Range	Reference
$E=15010\ p_{app}^{2.18}$	$p_{app} \leq 0.280$ g/cm^3	Morgan et al.
$E=6850\ p_{app}^{1.49}$	$p_{app} > 0.280$ g/cm^3	
$S_c=85.5\ p_{app}^{2.26}$	$p_{app} \leq 0.355$ g/cm^3	
$S_c=38.5\ p_{app}^{1.48}$	$p_{app} > 0.355$ g/cm^3	
$S_t=50.1\ p_{app}^{2.04}$	$p_{app} \leq 0.355$ g/cm^3	
$S_t=22.6\ p_{app}^{1.26}$	$p_{app} > 0.355$ g/cm^3	

2.5.3. Assignments of Loads and Boundary Conditions

The loads and boundary conditions were defined to simulate a fall on the greater trochanter (*Figure 12*) (Naylor, McCloskey et al. 2013). The impact force defined with respect to body weight and height of individual (see Appendix II), was applied to the greater trochanter. In the meantime, the femoral head and the distal shaft were constrained in one direction (i.e. x direction) and in both directions (i.e. x and y directions), respectively.

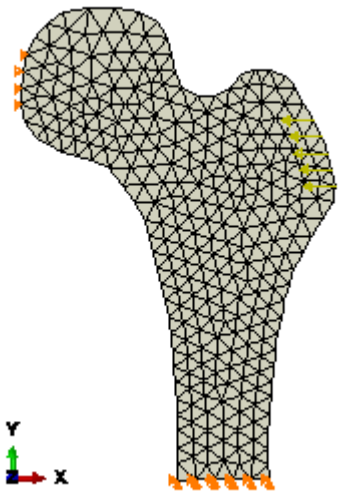

Figure 12: Force was applied on the greater trochanter. The femoral head was constrained in x direction, while the distal shaft was restricted in x-y directions.

2.5.4. Post Processing of FE analyses

To predict the proximal femur fracture load, stress based failure criterion (i.e. distortional energy that is also known as Von Mises stress) was used. A factor of failure was computed at the centroid of each element, in a similar way that had been presented by Keyak et al. (Keyak, Rossi et al. 1998):

$$\text{FOF} = \frac{\text{Von Mises Stress at element centroid (computed the FE model)}}{\text{Compressive strenght of element (derived from DXA scan)}} \quad (8)$$

The element which had a FOF value greater than one was assumed to be failed. Considering the fact that failure of few elements does not really represent fracture of the whole bone, the failure load was defined as the load which causes failure of contiguous elements with an area of 25 mm^2 (Naylor, McCloskey et al. 2013). To determine the failure load, the post-processing code was written in such a way that it starts searching for failed contiguous elements with an area of 25 mm^2 at the one-tenth of the impact force. If there is no such an area with the defined criterion, the FOF value was decreased with a factor β and the search

procedure started again. When the described criterion was satisfied, the failure load was calculated using the β value and taking the linearity of the developed FE model into consideration:

$$\text{Predicted Fracture Load} = \frac{\text{One tenth of the peak impact force}}{\beta} \quad (9)$$

2.6. Evaluation of FEA

To assess the effect of segmentation error in prediction of proximal femur fracture load, two different FE models were compared with each other. First, patient-specific FE model was developed based on known ground truth (i.e. proximal femur that had been manually segmented), while the second one was built up based on the proximal femur that had been segmented using AAM. During the evaluation process, 17 proximal femurs that had not been involved in the training phase of AAM and were used in the generation of FE models. A total of 34 FE analyses (17 analyses for manual segmentation and 17 analyses for segmentation using AAM) were performed. Based on the analyses, failure load was determined for each FE model (see Results section).

3. RESULTS

As mentioned, following the generation of AAM from a training set, an essential step is to evaluate the performance of segmentation. Based on the leave-one-out experiments, conducted with a set of 70 DXA scans, the mean point-to-curve error was found to be 1.2470 ± 0.6505 (mm) (with 95% confidence), over 70 DXA images. *Figure 13* depicts the frequency of errors for each experiment.

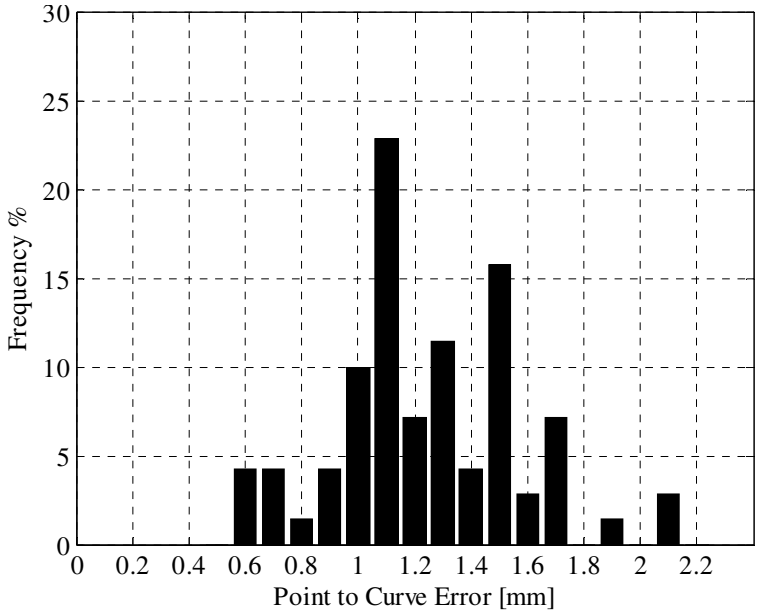

Figure 13: Frequency of point-to-curve errors in leave–one-out experiments.

Figure 14 presents the point-to-curve errors, distributed within the proximal femur. It should be noted that segmentation is less accurate at the femur head and at the greater trochanter compared to other regions.

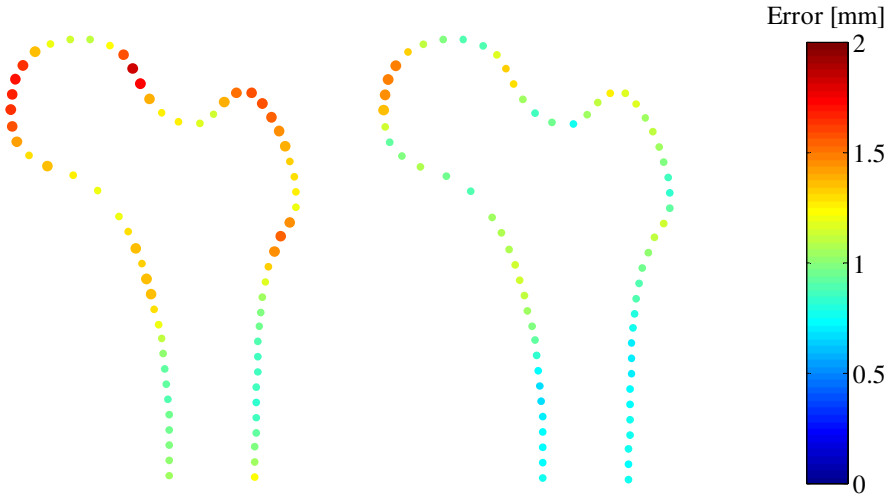

Figure 14: The mean (left) and a standard deviation (right) of error for each landmark.

Another essential step of the study was to conduct mesh convergence analysis to obtain mesh independent FE analyses results. In order to assess mesh density, Von Mises stresses (at element centroid) and average displacements were recorded at five elements for different mesh densities (see section meshing). As two particular cases (i.e. uniform and non-uniform

materials distribution within the proximal femur) were considered, mesh convergence study results are given for both cases. *Figure 15-16* and *Figure 17-18* represent Von Mises stress, displacements for uniform and non-uniform material distribution cases, respectively.

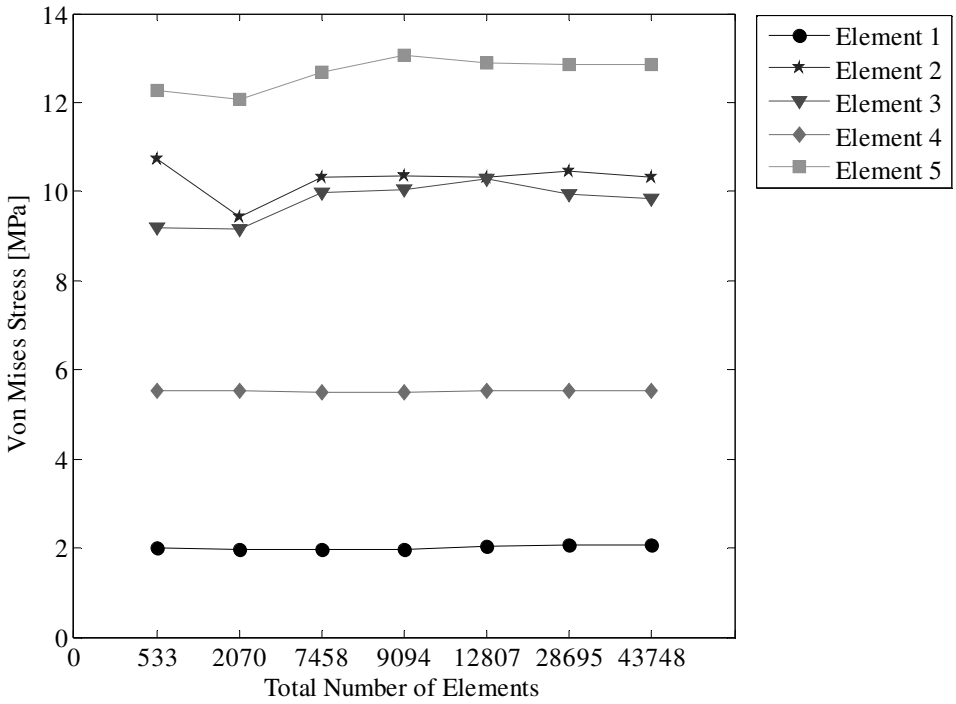

Figure 15: Von Mises Stress (MPa) at various mesh densities for uniform material distribution.

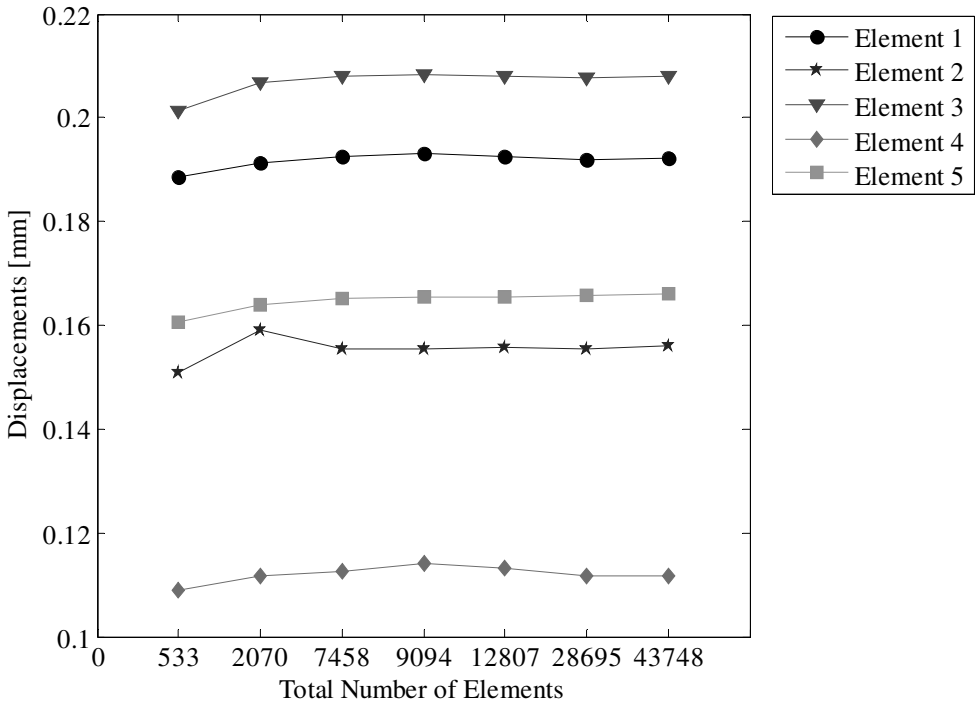

Figure 16: Displacements at various mesh densities for uniform material distribution.

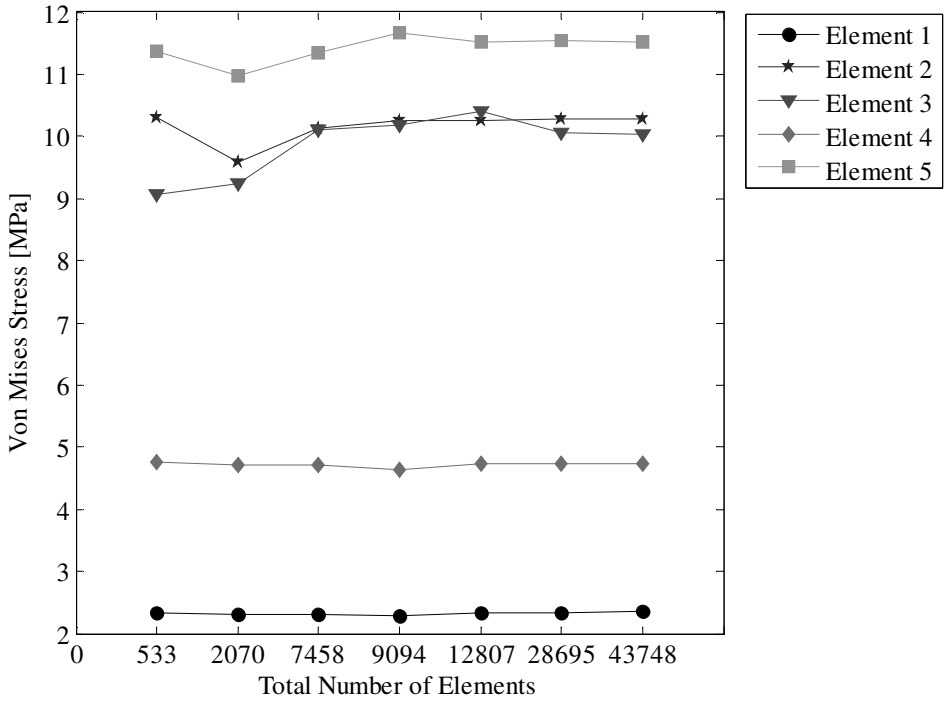

Figure 17: Von Mises Stress (MPa) at various mesh densities for non-uniform material distribution.

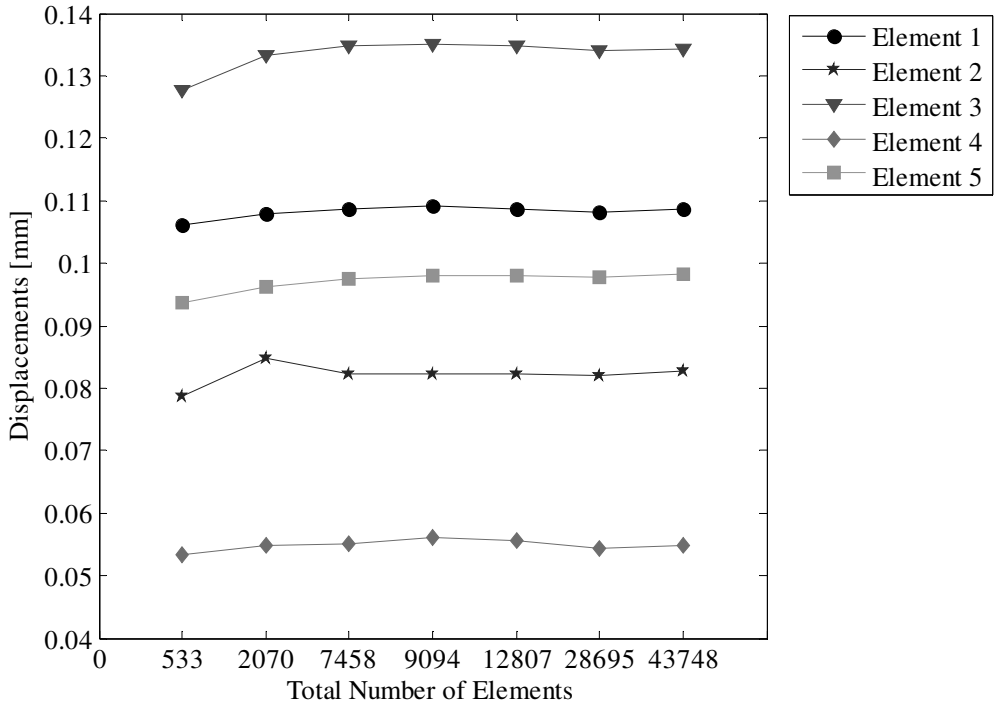

Figure 18: Displacements at various mesh densities for non-uniform material distribution.

Considering Von Mises stress results of uniform material distribution case, the relative error between the first (533 elements) and the second (2070 elements) mesh densities is 12%, while between the third (7458 elements) and the fourth (9094 elements) mesh densities is close to 4%. The relative errors for further mesh refinements are less than 4%. Furthermore, taking displacements for uniform material distribution into account, the relative error between the first two mesh densities is %5. The relative displacement error becomes less than 2% following the third mesh refinement. As in the Von Mises stress results of the case with uniform material distribution, the relative error remains under 4% after the third mesh refinement. Although relative displacement errors are moderately high compared to those of the uniform case, relative errors remain to be less than 3% after the third mesh refinement. Based on the relative errors and the time required for FE analyses, the sixth mesh density (28695 elements) was chosen and used throughout the study.

Furthermore, in order to evaluate the influence of segmentation error on the calculated proximal femur fracture load, point to curve errors were computed for each AAM-segmented 17 proximal femurs and compared with the known ground truth. Error was found to be 1.4169 ± 0.7499 (mm) (with 95% confidence) in all 17 proximal femurs. *Figure 19* visualizes the distribution of point-to-curve errors in each comparison.

Moreover, to evaluate the distribution of errors within the proximal femur, *Figure 20* presents the mean and a standard deviation of error for each landmark of the shape model. Based on this figure, as in the leave-one-out experiments, segmentation is seen to be less accurate at the femur head. On the other hand, considering the manual segmentations of 10 randomly chosen proximal femurs by three other students, point to curve error was found to be 1.0354 ± 0.4771 (mm).

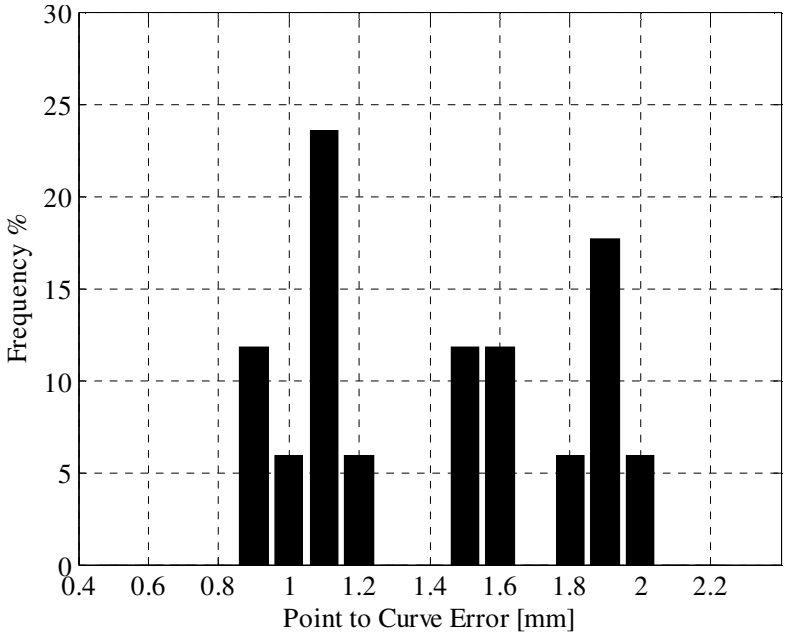

Figure 19: Frequency of point-to-curve errors for 17 proximal femur samples based on which FE models would be developed.

Figure 20: The mean (left) and a standard deviation (right) of error for each landmark.

Concerning the post-processing of the FE analyses, the proximal femur fracture load was estimated as 3870.9 ± 932.83 (N) and 3804.2 ± 850.11 (N) for 17 proximal femurs that had been manually segmented and for the same samples segmented using AAM, respectively. The relative error between two different segmentation cases was found to be 8.8 ± 6.5 (%). To investigate the correlation between the estimated fracture load in manual segmentation and in the case of using AAM, a simple linear regression on the failure loads F_{MAN} vs. F_{AAM} was performed (*Figure 21*). In figure, the coefficient of determination, R^2 together with the regression line equation are indicated. Based on the regression analysis, F_{AAM} shows a strong correlation (R^2 =0.8197) with F_{MAN}.

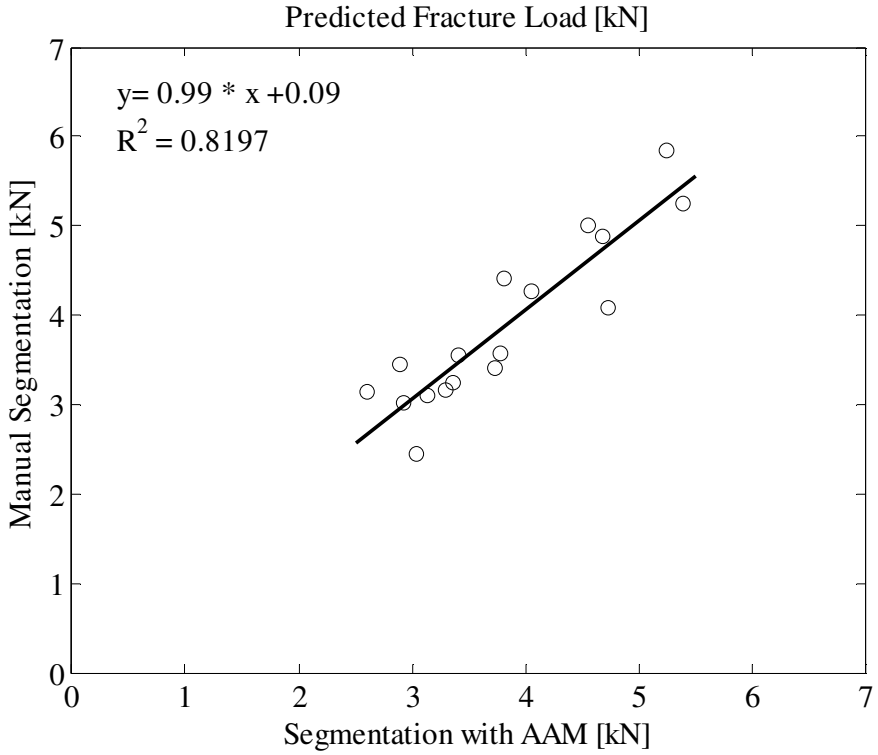

Figure 21: Predicted fracture load in manual segmentation, F_{MAN} vs. fracture load in the case of using AAM, F_{AAM}.

Moreover, *Figure 22* illustrates the distribution of Von Mises stress and the factor of failure (FOF) within one of the proximal femur samples that has relative failure load error close to the mean. These distributions were obtained at the predicted fracture load.

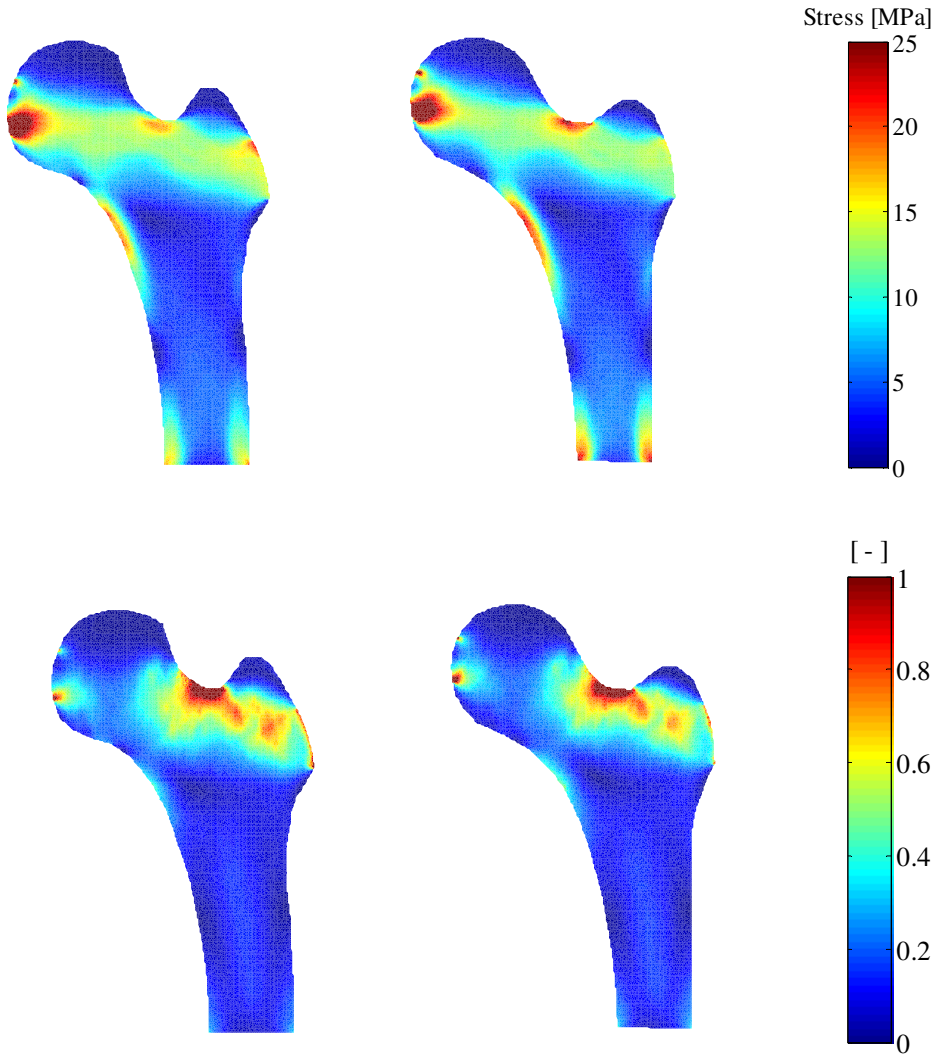

Figure 22: Distribution of Von Mises (top left, for manual segmentation; top right, for segmentation using AAM) and of FOF (bottom left, for manual segmentation; bottom right, for segmentation using AAM) within one of the proximal femur samples that has relative failure load error close to the mean.

Based on the FOF distributions, the most frequent failure areas (i.e. failed elements with a contiguous area of 25 mm^2) were found to be located at the superolateral part of the femoral neck. Furthermore, to find out whether the relative failure load errors are correlated with segmentation errors, correlation coefficient was computed. The scatter plot of segmentation error vs. relative failure load errors is given in *Figure 23.* No significant correlation was found (r=0.1414 and p=0.5883).

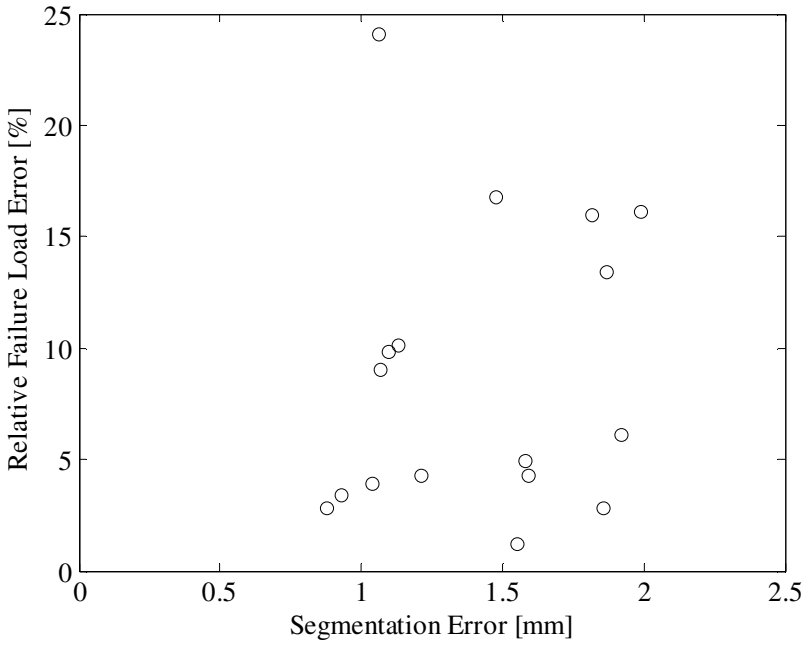

Figure 23: Relative fracture load errors vs. point-to-curve segmentation errors.

Moreover, point-to-curve errors were investigated considering four distinct regions of proximal femur: shaft, trochanter, neck and head. *Figure 24* depicts the *s*catter plots of relative failure load error in function of segmentation errors at the mentioned regions for all 17 proximal femurs. No correlation was observed between relative fracture load error and segmentation errors.

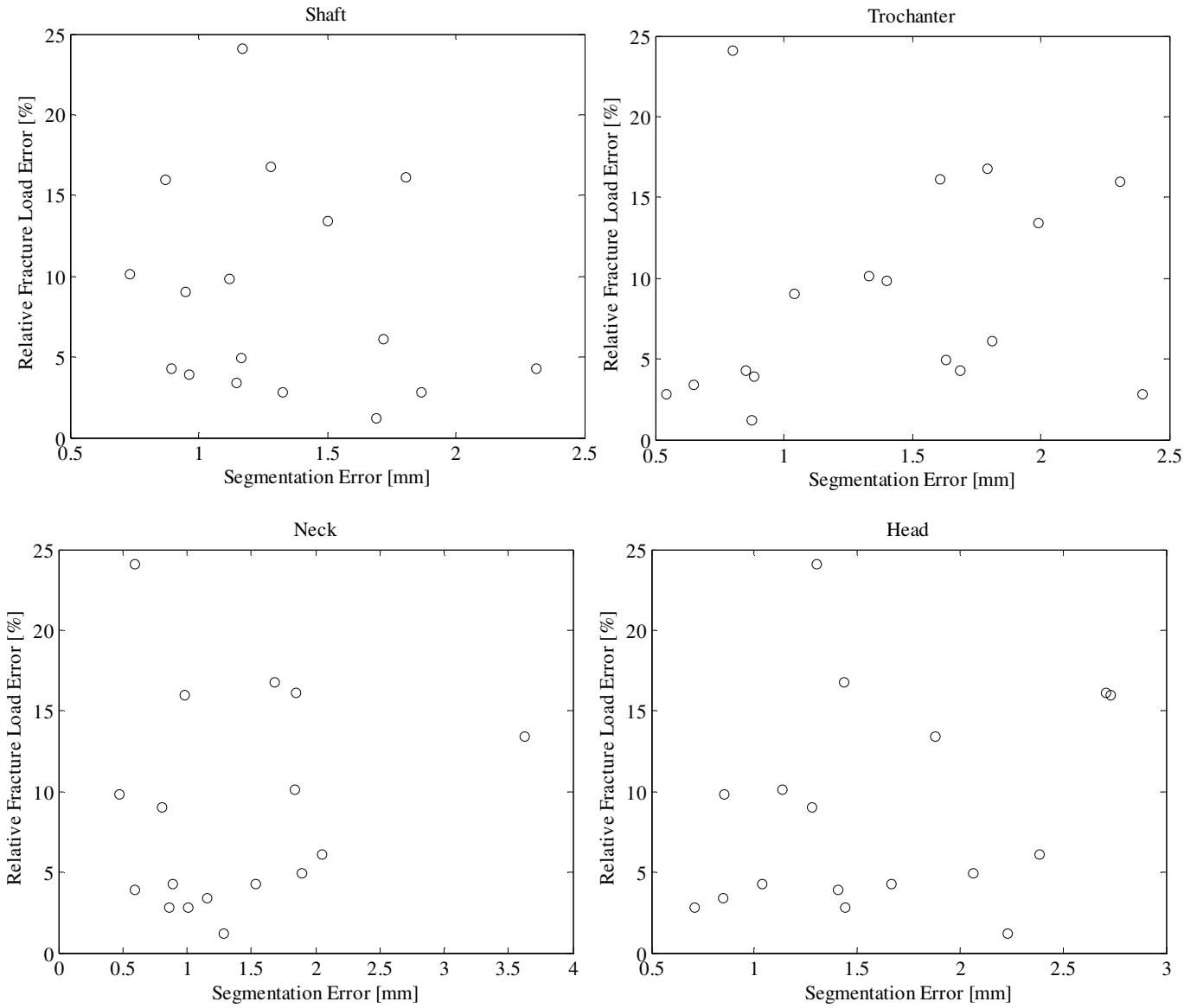

Figure 24: Relative fracture load errors vs. point-to-curve errors at Shaft (top left), at Trochanter (top right), at Neck (bottom left), at Head (bottom right).

4. DISCUSSION

As previously mentioned, use of DXA-measured aBMD is a common method to predict osteoporotic hip fractures in clinical settings. However, taking only the changes in aBMD into consideration is not enough to explain all low energy fractures (Vaananen, Isaksson et al. 2011). Therefore, it is essential to develop alternative methods that also consider the influence of other parameters (e.g. shape of the anatomical structure, load conditions), which are known to be associated with fracture. Development of FE models is a powerful tool to investigate the

strength of the bone in vivo and, thus, to estimate risk of fracture. Numerous FE models of proximal femurs had been generated based on QCT images (Cody, Gross et al. 1999, Keyak and Falkinstein 2003, Wirtz, Pandorf et al. 2003). However, QCT imaging technique is associated with high costs and high radiation doses for patients (Whitmarsh, Humbert et al. 2011). Therefore, it becomes essential to keep DXA as the standard imaging modality in the building of FE models. On the other hand, the main goal in development of alternative methods is to make them adaptable to clinical settings. Accordingly, the proposed methods need to be accurate and sufficiently fast for specific tasks, e.g. estimation of proximal femur fracture load. Within this study, the main aim was to develop a framework that covers the generation of subject specific FE models of proximal femurs based on DXA images. As FE model creation requires knowledge on the shape of the proximal femurs of individuals, segmentation process plays an important role for deriving the information directly from the images. Although Active Shape Models (ASM) is more widely used to search images to detect object of interest (Heimann and Meinzer 2009), within this study AAM approach was preferred for segmentation task. The main reason to choose AAM was to get familiar with this method, as it provides several benefits. Like ASM, following a proper initialization, segmentation of proximal femur using AAM can be realized within a few seconds. On the other hand, similar to ASM, AAM based framework can be used by any person who is not an expert in image analysis (Stegmann 2000). Moreover, as AAM can capture both the texture (i.e. intensities across the object of interest, thus distribution of bone densities) and the shape variations within a studied population and can be extended to 3D in straightforward manner (Kroon 2011), it is also be used in the reconstruction of 3D shape and spatial distribution of bone densities from DXA (Whitmarsh, Humbert et al. 2011). Afterwards, 3D patient-specific FE models, naturally more accurate in the prediction of proximal femur fracture loads compared to 2D FE models, can be generated based on reconstructed proximal femurs.

Leaving advantages and possible applications of AAM aside, performance of the AAM was evaluated by conducting leave-one-out experiments. Based on results, point-to-curve error was found to be 1.2470 ± 0.6505 (mm) (with 95% confidence). On the other hand, point-to-curve error in segmentation of 17 proximal femurs that were used in FE analyses was computed as 1.4169 ± 0.7499 (mm) (with 95% confidence). The segmentation errors were compared with those of the study conducted by Lindner et al. (Lindner, Thiagarajah et al. 2013), who built a fully automatic system based on the improved standard ASM in order to detect proximal femur in anteroposterior pelvic radiographs. The model was trained using 839 images and was tested on 266 images. Based on the results, mean point to curve error was found to be less than 0.9 mm (for 99% of images) and less than 3.3 mm (for 100% of images) (Lindner, Thiagarajah et al. 2013). Although the training sets and the methods used for segmentation purpose are different and it is not possible to compare the errors directly with each other, their proposed method is seen to be more accurate. However, referring to the maximum point-to-curve error, lower values (i.e. 2.1 mm for leave-one-out experiments and 2 mm in segmentation of 17 proximal femurs) were obtained compared to those (i.e. 3.3 mm) indicated in the study of Lindner et al. (Lindner, Thiagarajah et al. 2013). Additionally, to evaluate the performance of generated AAM in segmentation, inter-observer reliability in manual segmentation was tested. Based on the test, point-to-curve error was found to be 1.0354 ± 0.4771 (mm). Comparing this error with those of in 17 proximal femur segmentation, it can be indicated that AAM based segmentation is worthy to be improved and used for derivation of proximal femur from DXA.

One of the drawbacks of the created framework within this study is the need for user interaction to initialize the segmentation process. This might lead to have intra and inter operator errors. To handle this, Lindner et al. developed the fully automatic segmentation process (Lindner, Thiagarajah et al. 2013). Like in the study of Lindner et al., AAM based

segmentation procedure can be made fully automatic in future studies. Furthermore, as the effectiveness of statistical models comes from the modes of variation involved in the training data set, incorporation of more training examples in generation of the AAM might be useful to increase the accuracy of the proposed method. Moreover, to speed up the model generation phase and to improve the segmentation accuracy, canonical correlation analysis as described by Donner et al. can be used for construction of the derivative matrix *R* (Donner, Reiter et al. 2006, Heimann and Meinzer 2009). This approach had found to be more accurate in prediction of parameter updates compared to the numeric differentiation as it had been applied in this study. It had also been indicated that this method decreases the time consumed in training of AAM, by requiring fewer training examples.

Considering FE analyses results, estimated fracture loads for FE models which had been manually segmented and had been derived from DXA using AAM, were found to be 3870.9 ± 932.83 (N) and 3804.2 ± 850.11 (N) respectively. Referring to the study of Naylor et al., fracture loads were estimated to be 1820 N (ranging between 1265, 2648 N) for women with prevalent osteoporotic fractures and 2614 N (ranging between 1793, 3435) for women with non-fracture (Naylor, McCloskey et al. 2013). Furthermore, according to Amir et al., femoral strength had been indicated to be around 2500 N (2577.7 ± 1172.2 N) in women and 3200 N (3217.4 ± 1270.9 N) in men with prevalent osteoporotic fractures, while to be around 4000 N (3866.9 ± 1186.6 N) for women and 4600 N (4602.0 ± 1287.4) in men with non-fracture (Amin, Kopperdhal et al. 2011). Estimated proximal femur fracture loads, in this study, remains within the failure load ranges that were presented by Amin et al. (Amin, Kopperdhal et al. 2011). On the other hand, estimated fracture loads for both segmentation cases were higher compared to those of the study of Naylor et al. (Naylor, McCloskey et al. 2013). The difference might result in from the differences of data set used in the studies. In their study, (Naylor, McCloskey et al. 2013), only women with a mean age of 82 years (ranging between

75 and 95 years) were considered. However, within this study, although the sample size is small, the data set included men and women with an average age of 61.82 ± 10.7 years. As it had been indicated by Lochmüller et al., although no significant changes in bone strength were observed in men, the failure loads decreased in females with an increasing age (Lochmuller, Zeller et al. 1998). Furthermore, failure loads in men were found to be higher compared to those of in women (Lochmuller, Zeller et al. 1998, Amin, Kopperdhal et al. 2011). These studies might explain the root cause for the difference in estimated fracture loads with Naylor et al. (Naylor, McCloskey et al. 2013).

No correlation was observed between segmentation errors and relative failure load errors by taking the whole proximal femur and four different regions (i.e. neck, trochanter, shaft and head) into account. The resulting segmentation errors do not provide the necessary insight about the implications of the error on the geometry and thus, the strength of the femur. Therefore, it is believed that some specific geometrical measures (e.g. femoral neck axis length) would be beneficial in order to evaluate the effects of segmentation errors in prediction of proximal femur fracture load. Although no correlation was examined, it was noticed that even small errors (e.g. 1.06 mm) in segmentation process might result in larger errors (e.g. 24.1%) in prediction of fracture load. Even if the aim of the study and the methods used in this study are quite different from those of the study conducted by Väänänen et al. (Vaananen, Isaksson et al. 2011), the observations are in agreement. Väänänen et al. indicated that small errors in 3D shape reconstruction shape of proximal bone from a 2D BMD image leaded in larger errors in mechanical parameters (e.g. Von Mises stress). Considering this fact, it becomes essential to improve more the accuracy of the AAM based segmentation process, in future studies.

Within this study, during the definition and assignment of material properties, grouping of materials were not considered. However, defining material groups, thus reducing number of

material labels determined for each mesh element, the time required for FE analyses can be lessened. For material group definition, approach that had been described by Taddei et al. (Taddei, Pancanti et al. 2004) can be adapted.

In conclusion, created framework has room for improvement and is suitable to be used in clinical settings. It is relatively fast and requires user interaction only at the initialization of segmentation process. However, the initialization of search process might become error-prone and time consuming when AAM is applied to a large scale data. Therefore, it is better to make the framework fully automatic in this regard.

Moreover, before its use in clinical settings, additional work is needed to verify the performance of the developed method. With this purpose, a long term study would be useful which allows to estimate the proximal fracture load based on developed FE analyses and also to compare the outcome with clinical data. In future studies, large scale long term data set will be included in both developments of AAM, and of FE models.

Additionally, a different framework that aims for the generation of 3D patient specific FE models from DXA scans would be useful in order to increase the accuracy of predictions.

REFERENCES

Amin, S., D. L. Kopperdhal, L. J. Melton, S. J. Achenbach, T. M. Therneau, B. L. Riggs, T. M. Keaveny and S. Khosla (2011). "Association of hip strength estimates by finite-element analysis with fractures in women and men." Journal of Bone and Mineral Research 26(7): 1593-1600.

Arun, K. S., T. S. Huang and S. D. Blostein (1987). "Least-squares fitting of two 3-D point sets." Pattern Analysis and Machine Intelligence, IEEE Transactions on(5): 698-700.

Bryan, R., P. B. Nair and M. Taylor (2009). "Use of a statistical model of the whole femur in a large scale, multi-model study of femoral neck fracture risk." Journal of Biomechanics 42(13): 2171-2176.

Cody, D. D., G. J. Gross, F. J Hou, H. J. Spencer, S. A. Goldstein and D. P Fyhrie (1999). "Femoral strength is better predicted by finite element models than QCT and DXA." Journal of Biomechanics 32(10): 1013-1020.

Cootes, T. F., G. J. Edwards and C. J. Taylor (1998). "- Active appearance models." - 1407.

Cootes, T. F., G. J. Edwards and C. J. Taylor (2001). "Active appearance models." Pattern Analysis and Machine Intelligence, IEEE Transactions on 23(6): 681-685.

Cootes, T. F. and C. J. Taylor (2004). "Statistical models of appearance for computer vision." Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK March 8.

Donner, R., M. Reiter, G. Langs, P. Peloschek and H. Bischof (2006). "Fast active appearance model search using canonical correlation analysis." Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(10): 1690-1694.

Du, S., N. Zheng, S. Ying, Q. You and Y. Wu (2007). An extension of the ICP algorithm considering scale factor. Image Processing, 2007. ICIP 2007. IEEE International Conference on, IEEE.

Goodall, C. (1991). "Procrustes methods in the statistical analysis of shape." Journal of the Royal Statistical Society Series B-Methodological 53(2): 285-339.

Heimann, T. and H.-P. Meinzer (2009). "Statistical shape models for 3D medical image segmentation: A review." Medical Image Analysis 13(4): 543-563.

Jolliffe, I. (2005). Principal component analysis, Wiley Online Library.

Keyak, J. H. and Y. Falkinstein (2003). "Comparison of in situ and in vitro CT scan-based finite element model predictions of proximal femoral fracture load." Medical engineering & physics 25(9): 781-787.

Keyak, J. H., S. A. Rossi, K. A. Jones and H. B. Skinner (1998). "Prediction of femoral fracture load using automated finite element modeling." Journal of Biomechanics 31(2): 125-133.

Kroon, D.-J. (2011). Segmentation of the mandibular canal in cone-beam CT data, University of Twente.

Lindner, C., S. Thiagarajah, J. Wilkinson, G. Wallis and T. Cootes (2013). "Development of a fully automatic shape model matching (FASMM) system to derive statistical shape models from radiographs: Application to the accurate capture and global representation of proximal femur shape." Osteoarthritis and Cartilage.

Lochmuller, E. M., J. B. Zeller, D. Kaiser, F. Eckstein, J. Landgraf, R. Putz and R. Steldinger (1998). "Correlation of femoral and lumbar DXA and calcaneal ultrasound, measured in situ with intact soft tissues, with the in vitro failure loads of the proximal femur." Osteoporosis International 8(6): 591-598.

Morgan, E. F., H. H. Bayraktar and T. M. Keaveny (2003). "Trabecular bone modulus–density relationships depend on anatomic site." Journal of biomechanics 36(7): 897-904.

Naylor, K. E., E. V. McCloskey, R. Eastell and L. Yang (2013). "Use of DXA-based finite element analysis of the proximal femur in a longitudinal study of hip fracture." Journal of Bone and Mineral Research 28(5): 1014-1021.

Pegg, E. C., S. J. Mellon, G. Salmon, A. Alvand, H. Pandit, D. W. Murray and H. S. Gill (2012). "Improved radiograph measurement inter-observer reliability by use of statistical shape models." European Journal of Radiology 81(10): 2585-2591.

Poelert, S., E. Valstar, H. Weinans and A. A. Zadpoor (2013). "Patient-specific finite element modeling of bones." Proceedings of the Institution of Mechanical Engineers Part H-Journal of Engineering in Medicine 227(H4): 464-478.

Roberts, M. G., E. M. B. Pacheco, R. Mohankumar, T. F. Cootes and J. E. Adams (2010). "Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi-automatic segmentation." Osteoporosis International 21(12): 2037-2046.

Schileo, E., F. Taddei, L. Cristofolini and M. Viceconti (2008). "Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested in vitro." Journal of biomechanics 41(2): 356-367.

Scholz, R., F. Hoffmann, S. von Sachsen, W.-G. Drossel, C. Klöhn and C. Voigt (2013). "Validation of density–elasticity relationships for finite element modeling of human pelvic bone by modal analysis." Journal of Biomechanics.

Stegmann, M. B. (2000). "Active appearance models: Theory, extensions and cases." Informatics and Mathematical Modelling: 262.

Strom, O., F. Borgstrom, J. A. Kanis, J. Compston, C. Cooper, E. V. McCloskey and B. Jonsson (2011). "Osteoporosis: burden, health care provision and opportunities in the EU: a report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA)." Archives of osteoporosis 6(1-2): 59-155.

Taddei, F., A. Pancanti and M. Viceconti (2004). "An improved method for the automatic mapping of computed tomography numbers onto finite element models." Medical engineering & physics 26(1): 61-69.

Testi, D., M. Viceconti, A. Cappello and S. Gnudi (2002). "Prediction of hip fracture can be significantly improved by a single biomedical indicator." Annals of biomedical engineering 30(6): 801-807.

Vaananen, S. P., H. Isaksson, P. Julkunen, J. Sirola, H. Kroger and J. S. Jurvelin (2011). "Assessment of the 3-D shape and mechanics of the proximal femur using a shape template and a bone mineral density image." Biomechanics and Modeling in Mechanobiology 10(4): 529-538.

Vaananen, S. P., J. S. Jurvelin and H. Isaksson (2012). "Estimation of 3D shape, internal density and mechanics of proximal femur by combining bone mineral density images with shape and density templates." Biomechanics and Modeling in Mechanobiology 11(6): 791-800.

Westerweel, J. (1993). Digital particle image velocimetry: theory and application. Ph. D., Technical University of Delft.

Whitmarsh, T., K. D. Fritscher, L. Humbert, L. M. del Rio Barquero, T. Roth, C. Kammerlander, M. Blauth, R. Schubert and A. F. Frangi (2012). "Hip fracture discrimination from dual-energy X-ray absorptiometry by statistical model registration." Bone 51(5): 896-901.

Whitmarsh, T., L. Humbert, M. De Craene, L. M. Del Rio Barquero and A. F. Frangi (2011). "Reconstructing the 3D Shape and Bone Mineral Density Distribution of the Proximal Femur From Dual-Energy X-Ray Absorptiometry." Ieee Transactions on Medical Imaging 30(12): 2101-2114.

Wirtz, D. C., T. Pandorf, F. Portheine, K. Radermacher, N. Schiffers, A. Prescher, D. Weichert and F. U. Niethard (2003). "Concept and development of an orthotropic FE model of the proximal femur." Journal of Biomechanics 36(2): 289-293.