31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

An Automated Method to Segment the Femur for Osteoarthritis Research

Jeffrey W. Prescott, *Student Member, IEEE,* Michael Pennell, Thomas M. Best, Mark S. Swanson, *Student Member, IEEE,* Furqan Haq, Rebecca Jackson, Metin N. Gurcan, *Senior Member, IEEE*

***Abstract*—In this paper we develop a fully automated method for the segmentation of the femur in axial MR images and its use in the analysis of imaging biomarkers for osteoarthritis (OA). The proposed method is based on anatomical constraints implemented using morphological operations to extract the femur medulla and a level set evolution to extract the femur cortex. The average agreement of the automated segmentation algorithm with ground truth manual segmentations was 0.94 ± 0.03 calculated using the Zijdenbos similarity index (ZSI). A pooled variance t-test analysis found significant associations between the KL grade, a clinical measure of OA severity, and both the cross-sectional area (CSA) of the femur medulla (p = 0.02) and the ratio of the femur medulla CSA to the femur cortex CSA (p = 0.04) for women. No significant association between femur measurements and KL grade was found for men.**

I. Introduction

OSTEOARTHRITIS (OA) is a generally progressive joint disease characterized by substantial pain and disability. It is the most common form of arthritis [1], affecting up to 90% of the population over the age of 65 [2]. The causes of OA are known to be multifactorial (genetic, environmental, mechanical, biochemical, etc.), although characterization of these factors has been difficult beyond a qualitative framework. The identification of quantitative biomarkers for OA is an increasingly important objective, especially given the aging U.S. population and increased incidence of this burdensome disease. This paper explores the relationship between measures of the femur medulla and cortex on MRIs and a clinical measure of OA severity, the KL grade [3]. This paper also describes a completely automated method to segment and quantify the femur for biomarker analysis.

Known risk factors for OA include obesity, previous knee injury, and selected forms of physical activity – all of which can cause changes in mechanical loading across the knee joint [1]. Bone quality is known to play an intimate role in the pathogenesis of OA [4]. The restructuring of bone is a well known phenomenon caused by changes in mechanical stress [5]. Physiological compensation to handle the stress caused by joint changes in OA may therefore result in increased bone deposition and therefore increased bone size. Additionally, persons with radiographic evidence of OA have a higher bone mineral density (BMD) than those without OA [6]. Previous work has analyzed bone size and mechanics of the femur diaphysis [7]. In the current study, characteristics of the femur diaphysis that were analyzed were cross-sectional areas (CSA) and relative sizes of the medulla and cortex. The objective was to explore the relationship between these measures and clinical grades of OA severity.

Section II describes the image dataset used for the analysis; Section III presents the methodology of the automated segmentation and describes the statistical analysis performed using characteristics of the automatically segmented femurs; Section IV presents the segmentation results and statistical results; Section V discusses the implications of the results; and finally, Section VI offers conclusions and directions for future work.

II. Data

The data used for this analysis was from 103 subjects in the progression cohort of the Osteoarthritis Initiative's (OAI) public use dataset (www.oai.ucsf.edu). Population and OA characteristics of these subjects are summarized in Table I. T1-weighted axial scans of the thigh were acquired at 5 mm intervals in the range from 10 cm to 17 cm proximal to the medial femoral epiphysis of the right knee. In the current work, only the slice at 17 cm from the right thigh was analyzed. MRI acquisition parameters are listed in Table II.

Table I. Subject information.

Number of subjects		103
Age	Mean	61.2
	Standard deviation	10.2
Sex	Male	53
	Female	50
Ethnicity	Caucasian	85
	African American	16
	Asian	1
	Other non-white	1
KL grade	0	11
	1	19
	2	34
	3	35
	4	4

J. W. Prescott, M. S. Swanson, K. Powell and M. N. Gurcan are with the Department of Biomedical Informatics, Ohio State University, Columbus, OH 43210 (corresponding author J. W. Prescott phone: 614-292-4778; fax: 614-688-6600; e-mail: jeffrey.prescott@ osumc.edu).

M. Pennell is with the Division of Biostatistics, College of Public Health, Ohio State University Medical Center, Columbus, OH 43210.

F. Haq and T.M. Best are with the Department of Family Medicine, Ohio State University Medical Center, Columbus, OH 43210.

R. Jackson is with the Department of Endocrinology, Diabetes, and Metabolism, Ohio State University Medical Center, Columbus, OH 43210.

Table II. Parameters for acquisition of thigh MRIs.

Weighting	T1 Int
Plane	Axial
Number of slices	15
Field of view (mm)	500
Slice thickness (mm)	5
Skip (mm)	0
TE/TI (ms)	13
TR (ms)	600
X-resolution (mm)	0.977
Y-resolution (mm)	0.977

III. METHODOLOGY

A training set consisting of 30 subjects (15 males, 15 females) was randomly selected from the complete set of 103 subjects. All algorithm parameters were optimized for this training subset of patients. The remaining 73 subjects were used as a testing subset. Example images are shown in Fig. 1.

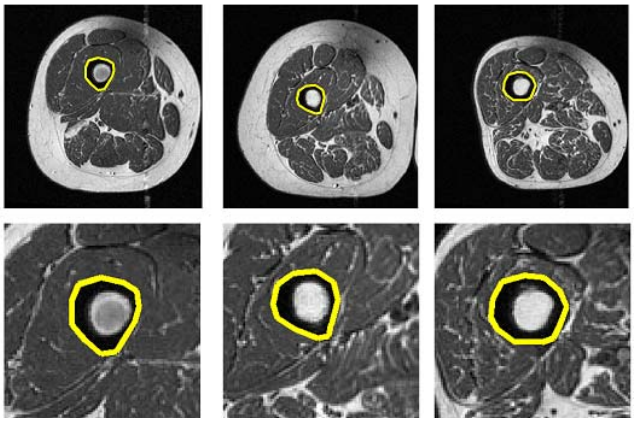

Figure 1. Images from three different subjects in the image dataset. Manual segmentations of the femur are shown in yellow. Top row displays full thigh images; bottom row displays enlarged region around femur. The femur medulla is the high intensity region in the center of the segmentations; the cortex is the dark region surrounding the medulla.

Images were first corrected for intensity bias fields due to magnetic field inhomogeneities. The N3 algorithm [8] was used, with parameters: full-width at half-maximum fwhm = 0.3, distance between basis functions d = 40 mm, noise term for deconvolution filter Z = 0.01. The resulting bias-field corrected images were then normalized to a range [0,1], with the top 0.05% of intensity outliers not included in the scale calculation. The order of the bias field correction followed by intensity normalization follows previous work by Madabhushi and Udupa [9].

The femur medulla was initially extracted by identifying pixels with intensity values above 0.5 in the scaled images with background pixels removed. This threshold was experimentally determined to provide good separation of fat and muscle. Morphological constraints were then applied to separate the femur medulla from regions corresponding to subcutaneous and intramuscular fat. The constraints used were:

1. The area of the region had to be between 30 and 500 pixels.
2. The pixels directly surrounding the area could contain no more than three high-intensity pixels (> 0.5) – this constraint used the anatomical characteristic that the femur medulla was always surrounded by a low intensity region corresponding to the femur cortex.
3. The solidity of the region (the proportion of pixels in the convex hull of the region that are also in the region [10]) had to be greater than 0.7. This constraint was required to ensure that there were no large areas of background pixels (intensity < 0.5) in the region.
4. An ellipse fitted to the region had to have a major axis to minor axis ratio less than 2. This ensured that the region was approximately circular.

The entire femur (including the cortex) was then segmented by evolving a level set with the boundary contour of the medulla used to define the initial level set function. It was found experimentally that by replacing the segmented femur medulla region with Gaussian noise of mean 0.05 and standard deviation 0.017 the level set contour would be more free to expand to the outer borders of the femur cortex, as there would not be a high value gradient edge corresponding to the femur medulla / cortex boundary. The level set evolution without reinitialization method described in [11] was used. The initial level set was created using the segmented femur medulla.

$$\phi_0(t) = \begin{cases} -1 & (x,y) \in \Omega_0 - \partial\Omega_0 \\ 0 & (x,y) \in \partial\Omega_0 \\ 1 & (x,y) \in \Omega - \Omega_0 \end{cases} \quad (1)$$

Where Ω was the domain of the level set function, Ω_0 was the area contained within the contour of the zero level set (i.e., pixels within the femur medulla), and $\partial\Omega_0$ was the zero level set (i.e., the boundary of the femur medulla). The evolution equation was

$$\frac{\partial\phi}{\partial t} = \mu\left[\Delta\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right] + \lambda\delta(\phi)div\left(g\frac{\nabla\phi}{|\nabla\phi|}\right) + \upsilon g\delta(\phi) \quad (2)$$

Where μ, λ, and υ were constant weighting coefficients for the terms in the evolution equation with experimentally determined values set to 0.05, 3, and -1.5; Δ was the Laplacian operator; *div* was the divergence function; and $\delta(\phi)$ was the Dirac delta of the level set function. The time step of the evolution was set to three. The edge indicator function, *g*, was calculated as

$$g = \frac{1}{1 + \left|\nabla G_\sigma * I\right|^2} \qquad (3)$$

Where G_σ was a 3 by 3 Gaussian smoothing kernel with standard deviation □ set to 0.5.

The evolution of the level set was terminated when the iterations hit a maximum value of 300. This value was experimentally determined to provide enough iterations for a complete femur segmentation.

The relationships of femur measurements to a clinical score of OA, known as the KL grade, were analyzed using a pooled variance t-test.

Image processing analyses were performed using Matlab, R2008a (Natick, MA). Statistical analyses were performed using JMP 8.0 (Cary, NC).

IV. RESULTS

Automated segmentation examples are shown in Fig. 2.

The performance of the training and test sets were validated by comparing the automated segmentations with manual segmentations performed by trained readers using ImageJ (rsbweb.nih.gov/ij/). The Zijdenbos similarity index was used to quantify the results.

$$ZSI = \frac{2*(A \cap M)}{|A| + |M|}$$

Where A was the binary mask of the automated segmentation and M was the binary mask of the manual segmentation. A ZSI value greater than 0.7 was considered to represent "excellent" agreement [12].

The training set ZSIs for automated vs. manual segmentations of the entire femur are shown in Table III, and the interreader agreement for the manual readers is shown in Table IV. In addition to the entire femur segmentations, Reader 1 also segmented the femur medulla manually. The ZSI for the automated segmentation of the medulla compared with the manual segmentation had a mean of 0.89 and a standard deviation of 0.06.

Manual segmentations of the entire femur of the testing dataset were performed by Reader 1 only. The mean ZSI between the automated segmentations and the manual segmentations of the testing dataset was 0.91 with a standard deviation of 0.16. However, in these 73 tests there were two complete failures of medulla segmentation and hence total femur extraction. For these two cases, the ZSI was 0. The ZSI mean and standard deviation with these two cases removed were 0.94 and 0.02, resp.

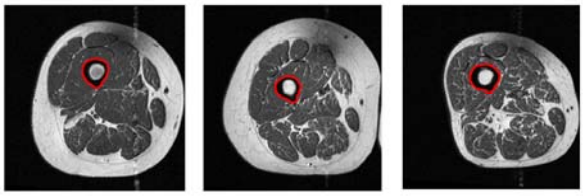

Figure 2. Results of the femur segmentation algorithm for three different subjects. Note the excellent agreement with the manual segmentations in Fig. 1.

Table III. Training set validation - ZSI means (standard deviations) for three trained readers' manual segmentations compared to automated segmentations.

Reader 1	Reader 2	Reader 3
0.94 (0.03)	0.94 (0.03)	0.94 (0.04)

Table IV. Interreader variability for manual segmentations of training dataset.

Reader 1 vs. Reader 2	Reader 1 vs. Reader 3	Reader 2 vs. Reader 3
0.96 (0.01)	0.97 (0.01)	0.97 (0.01)

One hundred and one subjects were used for the statistical analysis (the two subjects for whom the automated extraction failed were not included). The femur measurement data were examined for normality and equal variance and were found to satisfy each assumption. The relationship between each femur measurement and three potential confounders – sex, age, and BMI - was then analyzed. The only factor that was consistently related to each measurement was sex (Table V). Thus, the main analyses were stratified across the levels of this variable. Fig. 3 displays femur area vs. sex, with the center lines of the green diamonds corresponding to the sample means and the tips of the diamonds representing the 95% confidence intervals (CI).

Table V. Results (p-values) of the pooled variance t-test for femur CSA measurements and sex, age, and BMI. Age was grouped into two categories: < 62 and ≥ 62. BMI was group into two categories: < 30 and ≥ 30. * represents significance at p = 0.05 level.

	Stratified variable		
Femur Measurement	Sex	Age	BMI
Entire Femur CSA	< 0.0001*	0.19	0.02*
Medulla CSA	< 0.0001*	0.04*	0.16
Cortex CSA	< 0.0001*	0.86	0.02*
Medulla:Cortex CSA	0.02*	0.02*	0.64

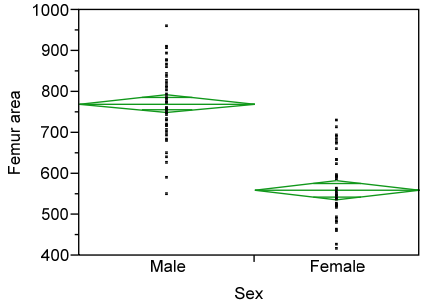

Figure 3. Plot of femur area samples vs. sex. Center lines of green diamonds correspond to means of samples, tips of diamonds represent 95% CI.

Pooled variance t-test analyses of femur medulla, femur cortex, and entire femur CSA and their associations with KL grade were performed, stratified by sex. The KL grades were combined into clinically meaningful groups: KL grades of 0-1 were combined into a single group (non-definitive radiographic OA) and KL grades of 2-4 were combined into a single group (definitive radiographic OA). TableVI shows the p-values from each pooled variance t-test. Significant associations at the p = 0.05 level were found between the KL grade and the medulla CSA and medulla:cortex CSA measurements for females (Table VII).

Table VI. Results (p-values) of the pooled variance t-test for femur CSA measurements and KL grade, stratified by subject sex. * represents significance at p = 0.05 level.

Femur Measurement	Male	Female
Entire Femur CSA	0.54	0.13
Medulla CSA	0.95	0.02*
Cortex CSA	0.34	0.77
Medulla:Cortex CSA	0.80	0.04*

Table VII. Mean (standard error) for femur measurements associated with KL grade for females.

	KL Grade	
Femur Measurement	0-1	2-4
Medulla CSA	102 (16)	145 (10)
Medulla:Cortex CSA	0.25 (0.04)	0.35 (0.02)

V. DISCUSSION

The performance of the automated segmentation algorithm was excellent, agreeing very closely with manual segmentations. In addition, the interreader agreement between the automated and manual segmentations was nearly equivalent to the interreader agreement between two manual readers.

There were two "misses" in the automated extraction of the medulla in the entire dataset. Both of these misses were from the test dataset. One was the result of an extremely low-intensity medulla, which was not entirely extracted by the initial thresholding operation and thus failed the solidity and circularity constraints. The other failure was due to a high-intensity, circular intramuscular fat region that was selected as the most likely region to be the femur medulla. These two failures may be overcome by using more statistical information about the location of the femur with respect to the entire thigh, and experiments with this type of constraint will be undertaken in the future.

A pooled-variance t-test analysis found a strong association between sex and each femur measurement. The pooled variance t-test analyses of KL grade and the femur measurements were therefore stratified by sex. For females, there was a significant association between femur medulla CSA (p = 0.02) and medulla:cortex CSA ratio (p = 0.04) with KL grade. There was no association between KL grade and any of the femur CSA measurements for males.

VI. CONCLUSION

An algorithm for automated segmentation of the femur in MRI scans was developed and validated. The algorithm separates the medulla from the cortex, allowing the morphological analysis of the different constituents of the femur and how these properties may be related to OA. Our analysis found that there were significant associations between femur CSA and sex. There were also strong relationships between femur medulla CSA and medulla:cortex CSA ratio with KL grade for females. Further work will incorporate measures of the entire available volume of the femur in the OAI studies. The automated femur segmentation method is currently being applied to the task of automated segmentation of the quadriceps muscles, which will be used in further research into imaging biomarkers of OA.

ACKNOWLEDGMENT

The authors would like to thank S. Mohan for his help with the manual segmentations of the femurs and K. Boussaid for comments on the initial draft of the manuscript. The project described was supported by Award Number R01LM010119 from the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

REFERENCES

[1] Felson DT, Lawrence RC, Dieppe PA, et al, "Osteoarthritis: New Insights. Part 1: The Disease and Its Risk Factors," *Ann Intern Med*, vol. 133, pp. 635-646, October 17. 2000.

[2] Buckwalter JA, Einhorn TA, Simon SR, and American Academy of Orthopaedic Surgeons., Orthopaedic Basic Science Biology and Biomechanics of the Musculoskeletal System. Rosemont, IL: American Academy of Orthopaedic Surgeons, 2000.

[3] Kellgren J, Lawrence J, "Radiological assessment of osteoarthritis." *Ann Rheum Dis* 1957; 16:494-501.

[4] Lajeunesse D, "The role of bone in the treatment of osteoarthritis." *Osteoarthritis and Cartilage*, (2004) 12, S34-S38.

[5] Martin RB, "Toward a unifying theory of bone remodeling." *Bone*, V. 26, No. 1, January 2000:1-6.

[6] Hochberg MC, Lethbridge-Cejku M, Tobin JD, "Bone mineral density and osteoarthritis: Data from the Baltimore Longitudinal Study of Aging." *Osteoarthritis and Cartilage*, (2004), 12, S45-S48.

[7] Stein MS, Thomas CDL, Feik SA, Wark JD, Clement JG, "Bone size and mechanics across the femoral diaphysis across age and sex." *Journal of Biomechanics*, 31 (1998), 1101-1110.

[8] Sled JG, Zijdenbox AP, Evans AC, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data." *IEEE Trans. Med. Imag.*, 17(1): 87-97, Feb 1998.

[9] Madabhushi A and Udupa JK, "Interplay Between Intensity Standardization and Inhomogeneity Correction in MR Image Processing." *IEEE Trans Med Img*, vol. 24, no. 5, May 2005.

[10] Gonzalez RC, Woods RE,. Digital Image Processing using Matlab, Prentice Hall. Upper Saddle River, NJ. 2003.

[11] Chunming L, Chenyang X, Changfeng G, Fox MD, "Level set evolution without re-initialization: a new variational formulation," *IEEE Conf. Comp. Vis. Patt. Recog.*, vol. 1, pp. 430-436, 20-25 June 2005.

[12] Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. "Morphometric analysis of white matter lesions in MR images: Method and validation." *IEEE Trans. Med. Imag.* 1994 Dec; 13: 716–24.