Automated Shape and Texture Analysis for Detection of Osteoarthritis from Radiographs of the Knee

Jessie Thomson[1,2], Terence O'Neill[2,3], David Felson[2,3,4], and Tim Cootes[1]

[1] Centre for Imaging Sciences, University of Manchester, UK
[2] NIHR Manchester Musculoskeletal BRU, Central Manchester NHS Foundation Trust, MAHSC, UK
[3] Arthritis Research UK Centre for Epidemiology, University of Manchester, UK
[4] Boston University, Boston, Massachusetts, USA

Abstract. Osteoarthritis (OA) is considered to be one of the leading causes of disability, however clinical detection relies heavily on subjective experience to condense the continuous features into discrete grades. We present a fully automated method to standardise the measurement of OA features in the knee used to diagnose disease grade. Our approach combines features derived from both bone shape (obtained from an automated bone segmentation system) and image texture in the tibia. A simple weighted sum of the outputs of two Random Forest classifiers (one trained on shape features, the other on texture features) is sufficient to improve performance over either method on its own. We also demonstrate that Random Forests trained on simple pixel ratio features are as effective as the best previously reported texture measures on this task. We demonstrate the performance of the system on 500 knee radiographs from the OAI study.

Keywords: Computer-aided diagnosis, Quantitative Image Analysis, X-ray Imaging, Imaging Biomarkers, Computer Vision.

1 Introduction

Osteoarthritis is considered to be one of the leading causes of disability today. It is a degenerative disease that effects the entire joint, degrading articular cartilage and deforming the surrounding bones and tissue of the affected joint. The disease is associated with pain, disability and substantial care costs each year [1]. Experienced clinicians currently perform the clinical grading of x-ray images. However, the features involved in OA are continuous, so the classification into the distinctive grades (normal, doubtful, minimal, moderate and severe) is often left to the subjective opinion of the grader. This quantisation and the uncertainties in assigning a grade make it hard to detect changes. Clinical trials thus require large numbers of subjects to identify effects of interventions reliably. There is a need for automated methods to make measurements and classifications to remove subjectivity. Work in the area of OA classification on radiographs is still limited.

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 127–134, 2015.
DOI: 10.1007/978-3-319-24571-3_16

The most significant approaches [2] and [4] include analysing textural information across the joint using image processing techniques. These methods look at the texture across the overall joint, with implicit shape information gathered from the radiographic images. However, from the pathophysiological properties of the progression of the disease, it is apparent that both shape and texture are useful for describing OA.

Recent work has shown quantification of the changes in texture and shape of Osteoarthritic knees. For instance [13] examines the correlation found between the Trabecular Bone fractal signature, seen in the texture under the tibial plateaus and the progression of OA [11]. [12] use statistical shape models to quantify changes in shape.

We describe a fully automated system which accurately identifies the outlines of the bones of the knee joint and combines both shape and texture features to better identify signs of disease. We use a Random Forest Regression Voting Constrained Local Model (RFCLM) (described comprehensively in [8]) to detect the position of the bones and to accurately locate a set of key landmarks across the tibia and femur. As well as giving a way of quantifying the shape of the bones this also allows us to define a consistent region of interest in the tibia in which to perform texture analysis. We train separate classifiers to distinguish between OA and non-OA, and demonstrate that a combination of the two independent measures leads to better overall discrimination.

2 Background

2.1 Fractal Signature

We use a method similar to that in [13]. This creates a set of features based on mean absolute pixel differences. The pixels are sampled using a circular region R_{xy}, where all pixels are compared against the central pixel (x, y). All pixels must lie at a Euclidean distance d_j between $4 \leq d_j \leq 16$ pixels (to reduce artefact errors). For each new pixel $(x_{ij}, y_{ij}) \in R_{xy}$, where $(x_{ij}, y_{ij}) \neq (x, y)$ and $i = 1, 2, ..., N_\theta$, $j = 1, 2, ..., N_d$, the absolute intensity difference is calculated $abs(I(x_{ij}, y_{ij}) - I(x, y))$, see Fig. (1), where N_θ is the number of angles considered and N_d is the max distance along the direction θ. These intensity differences are stored in a table $R(\theta, d)$ with θ corresponding to the direction, and d the distance. This is applied to each image patch over the region of interest. The differences at each direction and distance are summed and a mean calculated of the absolute differences across the texture region. This gives a representation of the texture changes at different angles across the region, a property found to be correlated to the progression of OA [11].

2.2 Statistical Shape Model

A linear statistical shape model [6] is learned from a set of aligned shapes by applying principal component analysis, leading to a model with the form

$$x = \hat{x} + Pb \tag{1}$$

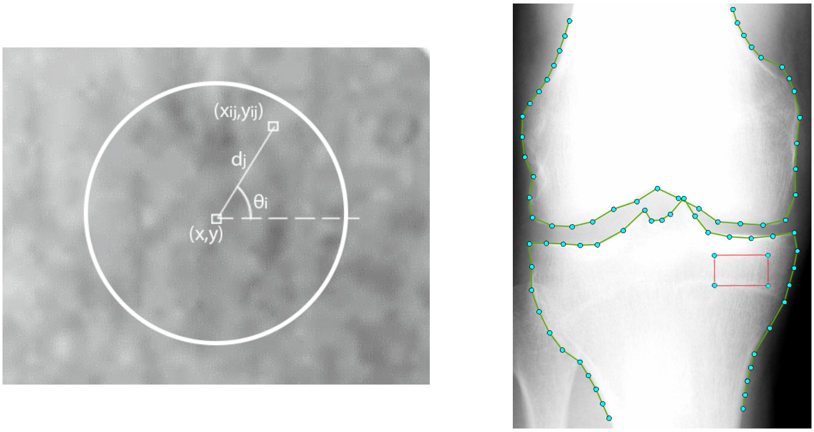

Fig. 1. Example of the circular region sampling to calculate Fractal Signature (left). Example of the 74 landmark points and 4 points computed from them outlining the ROI used for evaluating tibial texture (right).

where x is a vector representing the shape in a reference frame, $\hat{x}$ is the mean shape, P is the set of eigenvectors corresponding to the t highest eigenvalues, which describe different modes of variation, and b is a set of parameter values.

3 Method

3.1 Data

We use radiographic images from the OsteoArthritis Initiative [5], a longitudinal, prospective study of knee OA in the USA. The cohort has a collection of data from 4796 participants (men and women ages 45-79) taking scans annually across 8 years. The images have all been independently graded with Kellgren-Lawrence (KL) [7] grades 0 to 4, indicating the disease groups: normal, doubtful, minimal, moderate and severe. For the purpose of this study the grades have been split to distinguish either OA or non-OA. The non-OA class contains KL 0 and KL 1, whilst the OA class contains KL grades 2, 3 and 4. This distinction has been used in other OA classification studies, where KL1 is considered non-OA due to the features often described as doubtful. 500 AP knee images were selected from the cohort, based solely on the KL and Joint Space Narrowing (JSN) scores. The number of images within each grade are: KL0 - 110 (22%), KL1 - 142 (28.2%), KL2 - 87 (17.4%), KL3 - 118 (23.6%), KL4 - 43 (8.6%).

Knee Annotations: Each of the knees was annotated with 74 landmark points (Fig. 1). These annotations were then used to train a statistical shape model and an RFCLM object detection algorithm. The data was randomly split into two sets, one for training, one for evaluation. The manual annotations also provided an initial estimate of the OA classification performance to optimise training and analysis parameters.

3.2 Shape Model Matching

Using the set of 500 images a RFCLM model was trained and tested on halves of the data, producing shape model points on the test sets to use in the fully automated analysis. The RFCLM construction follows a similar process to that in [9], with a single global model to find the initial 2 points central to the joint, and 4 local RFCLM models built to iterate through increasing resolutions of the image, fitting the points to the best location at each stage. The accuracy of the algorithm was tested by comparing the detected point locations against the gold standard of the manually annotated points.

3.3 Classification Using Texture Information

The texture in regions of the tibia is known to change in knees with OA. We thus train two different classifiers to distinguish between the texture in knees with signs of OA and those without. As features we either use the data used to calculate fractal signature, or simple pixel ratios.

Region Selection: The region was defined under the medial tibial plateau. The analysis can be performed using both sides, however, to remove confounding effects from the fibula and due to the typically high prevalence of medial OA, only the side shown was analysed. This region was selected using two of the landmark points described in Sec. (3.1) along the sides of the tibia as guide points. Projecting a vector between the points a reference frame relative to the orientation, location and scale of the bone object can be made. From this the four points to make up the rectangle of the ROI can be found relative to the reference frame (see Fig. 1). The region within this rectangle is analysed to compute two different texture measures.

Fractal Signature: Using the method described in Sec. (2.1), we then made a slight alteration to the original method - collecting pixels outside of the set directions. For each pixel pair, the angle is rounded to the nearest fraction of the search region i.e. if 24 directions wanted; each angle is rounded to the nearest multiple of 15^o. This is to collect an average of pixel differences across a wider section of each search region. Due to time constraints in computation, the data used to originally compute Fractal Dimensions was omitted, instead all the data was used to train the Random Forest. The pixel samples were gathered from a region under the medial tibial plateau similar to that used in previous studies [11] and [13] that quantified fractal signatures of tibial texture.

Simple Pixel Features: As an alternative measure of texture we trained a Random Forest using pixel ratios as features in 32x32 pixel patches. For each image we randomly sampled 670 patches from the region on the tibia described above. A random forest was trained in which the decision at each node is a threshold on the ratio between intensities at two pixels from the patch. In this case when testing on new images we again extract 670 random 32x32 patches. We then compute the classifier response as the average of the RF response on each individual patch.

3.4 Shape Information

A Statistical Shape Model (described in section 2.2) was trained on a 74 point shape model, with 37 points per bone (tibia and femur). The shape model had 19 modes, sufficient to explain 90% of the variation in the training set. The shape parameter vectors, corresponding to b in Eq.(1), were then used to train a Random Forest classifier.

4 Results

4.1 Accuracy of Automatic Segmentation

We used two-fold cross validation (training on half the data, testing on the other half then swapping the sets) on 500 annotated images to evaluate the accuracy of the RFCLM when detecting the outline of the femur and tibia. Following [14] the mean point-to-curve distance was recorded as a percentage of a reference width across the tibial plateau, then converted to mm by assuming a mean width of 75mm. Results were: mean: 0.39% (0.29mm)$\pm 0.14mm$, median: 0.34% (0.26mm) and 95th percentile: 0.72% (0.54mm). These results are similar to those presented by [14], though on a different dataset.

4.2 Classification Experiments

To test the accuracy of each of the features, Random Forest [10] classifiers were trained on even splits of the 500 x-ray images taken from the OAI data. This classification was trained and tested on the clinically given KL grades, described in Sec. (3.1). The 500 images contained 244 OA and 256 non-OA. When split in half (250 train, 250 test images) the split to OA / non-OA was 120 /129 training, 127 / 122 testing.

4.3 Fractal Signature and Pixel Ratio Features

Fig. (2) shows the classification performance of the two texture based methods as ROC curves. Both methods have similar Area Under the Curve (AUC), with 0.74 for the Fractal Signature method and 0.75 for the pixel ratios. This suggests that the Random Forest is able to learn relevant features directly from the pixel ratios, avoiding the need to calculate more complex fractal texture signature. In the following we use the RF trained on pixel ratios to analyse the texture information.

Table 1. The AUC for each multi-feature analysis.

	Area Under ROC	
Analysis Method	Manual	Automated
Texture	0.745	0.754
Shape	0.796	0.789
Combined	0.844	0.845

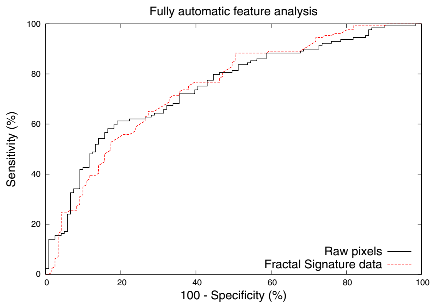

Fig. 2. The ROC curves for raw pixel ratio and Fractal Signature analysis.

4.4 Comparing Shape and Texture

Experiments were performed to compare the classifiers using (i) shape information, (ii) texture information (iii) both shape and texture.

Experiments were performed twice, once using the manually annotated landmark points (which define the shape and the region of interest for the texture analysis) and once using the results of the automated search (points for each image were generated by searching with a model trained on a different half of the data). The evaluation of each classifier was done by computing an ROC curve on the original train/test split, computing the AUC for each, and then finally an overall measure of accuracy was taken by using 5-fold cross-validation, repeated twice over all 500 images. The results for each stage can be seen in Table (1).

The two Random Forests trained independently on the separate features were then combined, with varying weights of the classification. The data was trained on two separate classifiers to allow more flexibility in the weight of classification from each model, and also to allow multiple texture samples per each image to be analysed. The optimal weighted combination of the two classifiers was found to be weighting by 2:1 in favour of the shape classification score, showing an

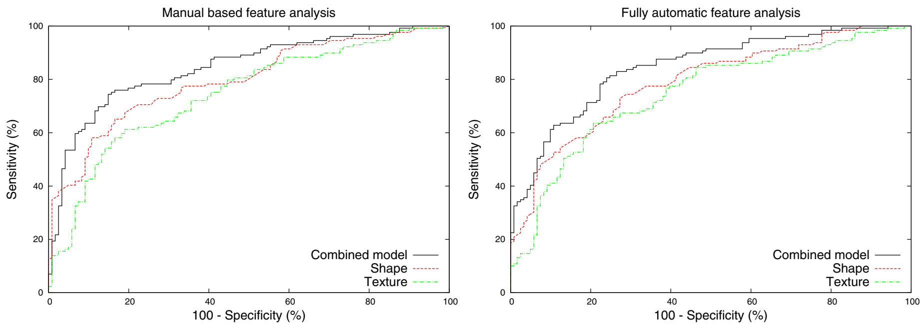

Fig. 3. ROC curves of classifiers using points from manual annotation (left), and fully automated approach (right)

accuracy of 0.84 in the fully automated annotations. The manual based and fully automated results for the 250/250 split in the data can be seen in Fig. (3).

5 Discussion and Conclusion

The experiments show that combining the results of the shape and texture based classifiers leads to a considerable improvement in overall classification performance, with an AUC of 0.849 compared to 0.789 for shape alone in the automated case.

The fully automatic system works as well as that based on manual annotation of the landmarks, due to the robustness and accuracy of the RFCLM matching algorithm. These results show that the combination of independently measured features of texture and shape create a stronger classifier when distinguishing Osteoarthritic knees. This study also looked at the classification comparison between Random Forests using the popular Fractal Signature and simple ratios pixel features. There is very little difference between the two when comparing AUC, suggesting that during training the Random Forest is learning appropriate combinations of pixel pairs to capture information encoded in the fractal signature in a different way.

Shamir *et al.* [2] reported obtaining classification rates of 86.1% distinguishing between OA and non-OA on a set of 350 images. They also distribute their OA classification software online (WND-CHARM). This algorithm was tested on the current dataset, achieving an AUC of 0.82, which is marginally less than our combined model. [3] focuses on using medial tibial texture to detect OA from a set of 102 images, also reporting a higher classification accuracy in comparison to WND-CHARM. Their model obtained a classification accuracy of 77.1%, which is higher than WND-CHARM (64.2%) and our own texture classification (69.2%). However, this was tested on a different dataset to that which we use.

In this paper we have focused on simple diagnosis (OA vs non-OA). In future work we will expand the classification to distinguish between separate KL grades, to determine the accuracy across each of the Osteoarthritic groups. The shape features were found to mainly separate the classes using the size of the joint space, in future work we will expand our method to analyse Osteophytes and bone remodelling. We will also investigate the effects of expanding the texture analysis to take more textural information from the images, from both the lateral tibial plateau, as well as sections of the femur, to determine whether including more information across both bones will increase the overall accuracy. Finally a regressor will be trained to give a continuous score (rather than a discrete value) to quantify the severity of disease. Since the process is completely automatic, it will be possible to apply the system to very large databases of images. We will examine the correlation between the results of our system and other clinical symptoms, and study how the response changes at different time-points as the disease progresses.

Acknowledgements. This report includes independent research funded by the National Institute for Health Research Biomedical Research Unit Funding Scheme. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Chen, A., Gupte, C., Akhtar, K., Smith, P., Cobb, J.: The global economic cost of Osteoarthritis: How the UK compares. Arthritis, vol. 2012 (2012)
2. Shamir, L., Ling, S.M., Scott, W.W., Bos, A., Orlov, N.: Knee X-ray image analysis method for automated detection of Osteoarthritis. IEEE Trans. Biomed. Eng. 56(2), 407–415 (2009)
3. Woloszynski, T., Podsiadlo, P., Stachowiak, G.W., Kurzynski, M.: A signature dissimilarity measure for trabecular bone texture in knee radiographs. Med. Phys. 37(5), 2030–2042 (2010)
4. Anifah, L., Purnama, I.K.E., Hariadi, M., Purnomo, M.H.: Osteoarthritis classification using self organizing map based on gabor kernel and contrast-limited adaptive histogram equalization. Open Biomed. Eng. J. 7, 18–28 (2013)
5. Lester, G.: Clinical research in OA. The NIH Osteoarthritis Initiative. J. Musculoskelet. Neuronal. Interact. 8(4), 313–314 (2008)
6. Cootes, T.F., Taylor, C.J.: Statistical Models of Appearance for Computer Vision. Technical report, University of Manchester (2004)
7. Kellgren, J.H., Lawrence, J.S.: Radiological assessment of Osteo-Arthrosis. Annals of the Rheumatic Diseases 16(4), 494–502 (1957)
8. Cootes, T.F., Ionita, M.C., Lindner, C., Sauer, P.: Robust and accurate shape model fitting using random forest regression voting. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part VII. LNCS, vol. 7578, pp. 278–291. Springer, Heidelberg (2012)
9. Lindner, C., Wilkinson, J.M., Consortium, T.A., Wallis, G.A., Cootes, T.F.: Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE Trans. on Med. Imaging 32(8), 1462–1472 (2013)
10. Breiman, L.: Random Forests. Machine Learning 45(1), 5–32 (2001)
11. Podsiadlo, P., Stachowiak, G.W.: Analysis of shape wear particles found in synovial joints. J. Orthop. Rheumatol. 8, 155–160 (1995)
12. Lee, H., Lee, J., Lin, M. C., Wu, C., Sun, Y.: Automatic assessment of knee Osteoarthritis parameters from two-dimensional X-ray images. In: First International Conference on ICICIC 2006, vol. 2, pp. 673–676 (2006)
13. Wolski, M., Podsiadlo, P., Stachowiak, G. W.: Directional fractal signature analysis of trabecular bone: Evaluation of different methods to detect early Osteoarthritis in knee radiographs. In: Proc. IMechE, vol. 223(2), Part H: J. Eng. Med. pp. 211–236 (2009)
14. Lindner, C., Thiagarajah, S., Wilkinson, J.M., arcOGEN Consortium, Wallis, G.A., Cootes, T.F.: Accurate bone segmentation in 2D radiographs using fully automatic shape model matching based on regression-voting. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 181–189. Springer, Heidelberg (2013)