Assessing hip osteoarthritis severity utilizing a probabilistic neural network based classification scheme

I. Boniatis a, L. Costaridou a, D. Cavouras b, I. Kalatzis b, E. Panagiotopoulos c, G. Panayiotakis a, *

a University of Patras, Faculty of Medicine, Department of Medical Physics, 26500 Patras, Greece
b Technological Educational Institute of Athens, Department of Medical Instrumentation Technology, 12210 Athens, Greece
c University of Patras, Faculty of Medicine, Department of Orthopaedics, 26500 Patras, Greece

Received 19 October 2005; received in revised form 24 February 2006; accepted 10 March 2006

Abstract

A computer-based classification system is proposed for the characterization of hips from pelvic radiographs as normal or osteoarthritic and for the discrimination among various grades of osteoarthritis (OA) severity. Pelvic radiographs of 18 patients with verified unilateral hip OA were evaluated by three experienced physicians, who assessed OA severity employing the Kellgren and Lawrence scale as: normal, mild/moderate and severe. Five run-length, 75 Laws’ and 5 novel textural features were extracted from the digitized radiographic images of each patient’s osteoarthritic and contralateral normal hip joint spaces (HJSs). Each one of the three sets of textural features (run-lengths, Laws’ and novel features) was separately utilized for assigning hips into the three OA severity categories, by means of a probabilistic neural network (PNN) classifier based hierarchical tree structure. The highest classification accuracy (100%) for characterizing hips as normal, mild/moderate or severe OA was obtained for the novel textural features set. Additionally, the novel textural features were used to design a mathematical regression model for providing a quantitative estimation of OA severity. Measured OA severity values, as expressed by HJS-narrowing, correlated highly (r = 0.85, p < 0.001) with the predicted values by the mathematical regression model. The proposed system may be valuable in OA-patient management.

Keywords: Hip; Osteoarthritis; Radiography; Texture; Classification

1. Introduction

Osteoarthritis (OA) is one of the most prevalent causes of disability worldwide. It is estimated that 30 million Europeans [1] and 40 million Americans [2] are affected by the disease. The condition is characterized by joint tissues alterations concerning both the articular cartilage and periarticular bone [3]. A variety of innovative imaging modalities can be employed in the clinical investigation of OA [4], while there is a growing interest regarding the use of MRI in the study of the disease, particularly in the knee joint [5]. However, in daily clinical routine, plain film radiography still remains the modality of first choice for assessing the existence and severity of OA. This is due to the fact that plain radiographs are easily accessible, readily available and economic. The radiographic hallmarks of hip OA comprise hip joint space narrowing (HJS-narrowing), subchondral bone sclerosis, formation of marginal osteophytes, development of femoral and acetabular subchondral cysts [6].

Qualitative as well as quantitative methods have been previously developed for the evaluation of OA severity and for monitoring of the disease progression. Qualitative methods refer to scoring systems, which are based on the subjective grading of OA radiographic features [7–9]. The Kellgren and Lawrence (KL) scale [7] is considered as the golden-standard for both cross-sectional and longitudinal epidemiologic studies, despite its deficiencies [10]. KL defines five categories of OA severity via an equal number of grades ranging between 0 and 4, with 0 indicating a normal hip joint and 4 a severe
osteoarthritic condition, while grades 1–3 define doubtful, mild and moderate OA, respectively [7]. According to the KL classification of OA, the radiographic features corresponding to each one of the grades of the KL scale are: grade 0, no features; grade 1, possible narrowing, possible osteophytes; grade 2, definite HJS-narrowing, definite osteophytes and slight sclerosis; grade 3, marked HJS-narrowing, moderate osteophytes, some sclerosis and possible cysts formation, possible deformity of femoral head and acetabulum and grade 4, large osteophytes, gross HJS-narrowing, severe sclerosis, cysts and marked deformity.

Quantitative assessment of hip OA refers to measurements of the HJS-width and/or the HJS-area on serial radiographs in order to monitor the progression of the disease in the same patient [11,12]. On the other hand, for classifying a hip as normal or osteoarthritic, thresholds have been introduced concerning the minimum hip joint space width [13,14], as measured on a single radiograph. However, to our knowledge, there have been no studies for quantifying the severity of the disease from a single radiograph, so that it can be applied to the general population.

In a digital image, texture is characterized by tonal and structural properties. Tonal properties refer to the intensities of the image pixels, while structural properties are related to the spatial organization of the image pixels. Thus, texture depicts the spatial distribution of pixel intensities in the image [15,16].

Texture analysis refers to algorithms developed to quantify image texture information that may, or may not, be perceived visually. Considering that medical images visualize various aspects of biological structures, texture analysis of digitized medical images can provide useful information relevant to the structure and status of biological tissues [17].

Regarding digitized radiographs, pixel intensities are related to X-ray attenuation, and thus the spatial distribution of gray levels is the result of projecting on two-dimensional level three-dimensional structures [17,18]. The radiographic HJS is a two-dimensional projection of the superimposed three-dimensional anatomical structures of articular cartilage, posterior acetabular wall and iliac bone. Articular cartilage consists of chondrocytes and a supportive matrix, labelled as the extracellular matrix. The main constituents of extracellular matrix are: water, collagen and proteoglycans. The mechanical properties of articular cartilage are closely related to the biochemical composition, as well as to the structure organization of extracellular matrix components [1]. Collagen fibers are associated to the tensile strength of cartilage, while the proteoglycans are correlated to cartilage elasticity [19]. Due to its strength and elasticity, the articular cartilage provides a bearing surface with low friction, facilitating the distribution of loads between adjacent bone surfaces in synovial joints [20]. In the osteoarthritic condition, degeneration of the articular cartilage differentiates its biomechanical properties causing the remodelling of subchondral bone, due to increased joint loading [18]. In addition, alterations concerning the structure, as well as the mechanical properties of subchondral bone have been associated to cartilage degeneration in OA [21,22]. These joint tissue alterations are expected to affect the radiographic texture of the HJS region. Thus, textural features extracted from HJS-ROIs, are expected to provide valuable information regarding the condition of anatomical structures in the hip joint. Texture based classification schemes have been previously implemented in a variety of medical imaging applications [23], while radiographic texture has been investigated as an alternative to the quantitative assessment of OA severity in case of knee joint [18].

In previous studies, neural network based classification schemes have been utilized for diagnosing osteoporosis and osteoarthritis employing stained histological images of the trabecular bone [24], for assessing defective lumbar vertebras from radiographic images [25], and for discriminating normal from abnormal knee joints [26]. However, to the best of our knowledge, a texture based pattern recognition approach has not been referred to the literature neither for the classification of hips into OA severity categories nor for the quantitative assessment of the disease from radiographic images of the HJS.

In this study, a computer-based image analysis system was developed for characterizing hips either as normal or as of mild/moderate or severe OA, by employing texture analysis methods on digitized pelvic radiographs. Additionally, the system was designed so as to quantify the severity of the hip joint OA. The system is intended to act as a diagnosis decision support tool to un-experienced physicians and/or as a second opinion tool to experienced ones.

2. Materials and methods

2.1. Patients and examination protocol

Thirty-six hips (18 normal and 18 osteoarthritic), corresponding to 18 patients with verified unilateral hip OA were included in this study. Patients’ ages ranged between 49 and 83 years (mean age ± S.D.: 65.6 ± 9.9 years). The American College of Rheumatology criteria [27] were used for OA diagnosis. The latter was based on clinical and radiographic assessment. More specifically, a hip was characterized as osteoarthritic if pain (associated to hip joint use) was reported in combination with the presence of osteophytes (femoral or acetabular) and/or joint space narrowing on pelvic radiographs.

Despite the fact that a digital radiography system would render the film digitization (necessary for subsequent texture analysis) obsolete, a conventional screen–film system was used for obtaining pelvic radiographs. This selection was based on the fact that the proposed classification approach was intended to be utilized in combination with conventional screen–film systems, which are of higher availability in daily
clinical routine in comparison to the digital radiography systems (limited availability, due to high economic cost).

All radiographs were obtained following a specific radiographic protocol, which comprised use of the same X-ray unit (Siemens, Polydoros 50, Erlangen, Germany), tube voltage 70–80 kVp, 100 cm focus to film distance, alignment of the X-ray beam 2 cm above the pubic symphysis, and use of a fast screen and film cassette (30 cm × 40 cm). Radiographs were digitized employing a laser digitizer suitable for medical applications (Lumiscan 75, Lumisys, Sunnyvale, CA, USA) [28]. In order to ensure the reliability of the digitization process, the digitizer performance was evaluated employing a quality control protocol [29]. Images were digitized to 12 bits (4096 gray levels) and 146 ppi (0.17 mm pixel size).

2.2. Patient grouping

For the needs of the present study, hips were grouped into three OA severity classes by three experienced orthopaedists, who assessed hip OA employing the KL severity scale [7]. Each of the orthopaedists graded OA by assigning a KL grade to the examined hips. In order to establish a golden-standard, only those exams of common consent were retained for the purposes of the present study. Accordingly, normal hips were assigned to normal/doubtful (KL = 0, 1) class, hips with a KL score of 2 or 3 were classified as mild/moderate, and hips exhibiting severe radiographic signs (KL = 4) were assigned to the severe class.

2.3. Delineation of radiographic hip joint space

On each pelvic radiograph, two HJS regions of interest (ROIs), corresponding to osteoarthritic and contralateral-normal hips, were determined, employing custom developed software (Fig. 1). Specifically, an algorithm realizing the Contrast-Limited Adaptive Histogram Equalization (CLAHE) method [30] was implemented in Matlab (The MathWorks Inc., Natick, USA) to enhance the digitized radiographs and, thus, to emphasize the articular margins of the hip joint. An acute angle of 45°, encompassing the weight-bearing portion of the hip joint, provided the medial and lateral limits of the HJS-ROI [11]. The medial limit was defined by the line joining the centre of femoral head (summit of the angle) and the highest point of the homolateral sacral wing and the lateral limit drawn at 45° to the medial limit (Fig. 1). Within this angle, the operator delineated manually the articular margins of the joint (edge of the femoral head and inferior margin of the acetabulem) using a graphics cursor. The original gray level values of the delineated HJS-ROI (see Fig. 2) were subjected to further analysis for the extraction of textural features. Texture analysis could not be performed in cases characterized by completely loss of the HJS due to narrowing, so an acceptance threshold was set (number of pixels corresponding to the ROI greater than 200).

Fig. 1. Determination of the hip joint space ROI. kOm: acute angle of 45° defined by the patient’s standard anatomical landmarks encompassing the examined ROI. Dotted lines represent the delineated by the physician articular margins.

Fig. 2. Example of segmented hip joint space ROI corresponding to Fig. 1.
2.4. Textural features generation

Textural features were generated from the segmented HJS-ROIs by the following methods: Gray Level Run Length Matrices (GLRLM) [31] and Laws’ Texture Energy Measures [32]. Additionally, novel textural features labelled as “Texture Energy Run Length” features were calculated.

2.4.1. Gray Level Run Length Matrices features

The GLRLM provides information related to the spatial distribution of gray level runs (i.e. pixel-structures of same pixel value) within the image. Textural features extracted from GLRLM evaluate the distribution of small (short runs) or large (long runs) organized structures within the HJS-ROI. From each HJS-ROI, five run-length features were generated and four values were computed for each feature, corresponding to the angles of 0, 45, 90 and 135°. In order to safeguard against rotation variability, the feature mean value was used [31].

2.4.2. Laws’ Texture Energy Measures features

According to the method proposed by Laws, textural features were extracted from images that had been previously filtered by each one of the 25 Laws’ masks or kernels [32]. These filtered images were characterized as Texture Energy images (TE_images). Averaging the filtered images corresponding to symmetrical kernels (such as $E_kL_K$ and $L_KE_k$), and taking into account that 20 out of 25 kernels are symmetrical one to each other, 15 TE_images were produced. From each 1 of the 15 TE_images, 5 first-order statistics [33] (mean, standard deviation, range, skewness and kurtosis) were computed, giving in total 75 Laws’ textural features: 5 sets of 15 features each, with each 15 feature set corresponding to each one of the 5 first-order statistics, i.e. 15 features for the mean value, 15 features for the standard deviation, 15 for the skewness, 15 for the kurtosis and 15 for the range.

2.4.3. Texture Energy Run Length features

In the present study, new features were proposed, based on combination of GLRLM features and Laws’ Texture Energy Measures, aiming to further improve the discriminatory power of these features. Specifically, five run-length features [31] were extracted from the TE_images [32]. The proposed textural features were labelled as “Texture Energy Run Length” features. The suggested approach comprised the following steps:

(i) The HJS-ROI image was convolved with the $L_KL_K$ Laws’ kernel:

$$F(x, y) = \sum_{s=-m}^{+m} \sum_{t=-m}^{+m} h(s, t) I(x + s, y + t),$$

$$m = \frac{k - 1}{2}, \quad k = 3, 5 \text{ or } 7$$

where $F(x, y)$ is the resulting filtered image, $I(x, y)$ the input HJS-ROI image and $h(s, t)$ is the $L_KL_K$ kernel [32]. In subsequent analysis, the $L_5L_5$ kernel, shown below, was used:

$$h(s, t) = \begin{bmatrix}
1 & 4 & 6 & 4 & 1 \\
4 & 16 & 24 & 16 & 4 \\
6 & 24 & 36 & 24 & 6 \\
4 & 16 & 24 & 16 & 4 \\
1 & 4 & 6 & 4 & 1
\end{bmatrix}$$

(ii) The TE_image of the HJS-ROI was formed by replacing every pixel in the $F(x, y)$ with a Texture Energy Measure (TEM) associated to the pixel, according to the scheme:

$$\text{TE_image}(x, y) = \sum_{j=-(l-1)/2}^{+(l-1)/2} \sum_{i=-m}^{+m} |F(x + s, y + t)|,$$

where $m = (l - 1)/2$, and $l = 3, 5, \ldots, 15$.

(iii) The GLRLM along direction $\theta$ of the TE_image was computed, as described by Galloway [31] and the following five Texture Energy Run Length (TERL) features were extracted from the relations:

$$\text{TERL}_1 = \frac{1}{P} \sum_{j=1}^{R} r_{\theta}(j)$$

$$\text{TERL}_2 = \frac{1}{P} \sum_{j=1}^{R} r_{\theta}(j)^2$$

$$\text{TERL}_3 = \frac{1}{P} \sum_{i=0}^{G-1} |g_{\theta}(i)|^2$$

$$\text{TERL}_4 = \frac{1}{P} \sum_{j=1}^{R} |r_{\theta}(j)|^2$$

$$\text{TERL}_5 = \frac{1}{PN} \sum_{j=1}^{R} r_{\theta}(j)$$

where $g_{\theta}(i, j)$ represents each element of the GLRLM computed along direction $\theta$, $j$ represents the length of the run for gray tone $i$, $G$ and $R$ are numbers of tones and run-lengths in the TE_image, respectively, $PN$ is the number of pixels in the TE_image, while the $r_{\theta}$, $g_{\theta}$ and $P$ are defined in the relations:

$$r_{\theta}(j) = \sum_{i=0}^{G-1} q_{\theta}(i, j)$$

$$g_{\theta}(i) = \sum_{j=1}^{R} q_{\theta}(i, j)$$

$$P = \sum_{i=0}^{G-1} \sum_{j=1}^{R} q_{\theta}(i, j) = \sum_{i=0}^{G-1} g_{\theta}(i) = \sum_{j=1}^{R} r_{\theta}(j).$$

Four values were computed for each feature, corresponding to the angles of 0, 45, 90 and 135°. In order to safeguard...
against rotation variability, the feature mean value was used [31].

Finally, all textural features employed in the present study were normalized to zero mean and unit standard deviation [34], according to the relation:

\[ x_{i, \text{norm}} = \frac{x_i - \mu}{\sigma} \]  

(5)

where \(x_{i, \text{norm}}\) is the normalized value of the \(x_i\) textural feature, while \(\mu\) and \(\sigma\) are the mean value and standard deviation, respectively, of feature \(x_i\) over all HJS-ROIs.

2.5. Classification of hips into OA severity categories

A probabilistic neural network (PNN) based hierarchical tree structure was developed for the classification of hips into three OA categories labelled as normal, mild/moderate and severe. PNN is practically a combination of neural networks with statistical theory, since it encompasses both the Bayes’ classification approach (theorem for conditional probability) and the Parzen’s estimators of probability density functions [35]. Fig. 3 shows the PNN based structure employed in the present study. As it may be observed, the classification scheme comprised two levels. In the first level, the discrimination between normal and osteoarthritic hips was performed. In the second level, the hips that had been characterized by the system as osteoarthritic were further classified as of mild/moderate or of severe OA.

At each level of the hierarchical tree structure, the PNN discriminated input vectors, formed each time by each of the three sets (run-length, Laws’ and TERL) of textural features extracted from each HJS-ROI, by means of the equation:

\[ g_j(z) = \frac{1}{(2\pi)^{p/2}\sigma^pN_j} \sum_{i=1}^{N_j} e^{-\|z - z_i\|^2/2\sigma^2} \]  

(6)

where \(p\) is the dimensionality (number of features) of the input pattern, \(\sigma\) is a smoothing parameter, \(z_i\) the \(i\)th training input pattern, \(z\) the unknown pattern to be classified and \(N_j\) is the number of patterns forming the class \(j\) [35].

Regarding the PNN structure, it is a feed-forward neural network, characterized by a high degree of parallelism. As presented in Fig. 3, the basic PNN architecture consists
of an input layer, a pattern layer, a summation layer and an output layer. The input layer stores temporarily each pattern vector, which is fed to the network. The number of neurons (nodes) that structure the input layer is equal to the dimensionality (p) of the input pattern. For input patterns formed by each one of the individual feature sets, the number of nodes of the input layer was 5 for the run-length, 15 for the Laws’ and 5 for the TERL features, respectively. Each input pattern is mapped to each one of the neurons of the pattern layer. Each neuron in the pattern layer represents a training pattern. In the first level of the hierarchical tree structure, the pattern layer for the normal as well as for the pathological class comprised 18 neurons. In the second level, the pattern layer for the mild/moderate class comprised nine nodes, while the same number of neurons was used for the severe class. In the pattern layer, the Euclidean distance between the input and each training pattern is computed. The activation function (represented by the exponential term in Eq. (6)) is then applied to provide the output of the pattern neuron. The summation layer has one neuron for each class, and implements the summation term of Eq. (6) for the outputs of the patterns corresponding to the class. As it can be observed from Fig. 3, each summation neuron is connected to the neurons of the corresponding pattern layer. The output layer contains one neuron and assigns the input vector to a class by implementing a classification rule. In particular, the unknown pattern is classified to the class with the highest discriminant value of $g_j(z)$. PNN’s main advantages are that it is fast to train and that data normality is not a prerequisite. In the training process, neither iterative procedures are used, nor feedback paths are required, since the PNN is a feed-forward and one pass structure [35].

At each level of the hierarchical tree structure, the PNN classifier was trained separately by each one of previously described individual sets of textural features. Within this context, a comparative study concerning the discriminatory power of each one of the sets of textural features was performed. The performance of the proposed classification scheme was evaluated in terms of sensitivity, specificity and overall accuracy. The classifier was validated employing the Leave-One-Out (LOO) method, i.e. the classifier was designed by all but one sample of the data set, which was then classified to one of two classes [34]. In order to determine the optimum feature combination for each classification task (i.e. the combination providing the highest overall classification accuracy with the minimum number of features for each feature set), an exhaustive search procedure was followed. In particular, the classifier was designed by means of every possible feature combination (i.e. 2–5 feature combinations), each time testing the classifier’s performance and finally selecting that feature combination that demonstrated the highest classification accuracy with the smallest number of features [34]. The exhaustive search procedure was performed for each one of the individual sets of textural features employed in the study.

2.6. Quantification of hip OA severity

The TERL features that were extracted from the osteoarthritic HJS-ROIs were also exploited in quantifying OA severity in the patients, since the utilization of the specific features resulted in the highest classification accuracies (see Tables 1, 2 and 5 and Figs. 5 and 7).

First, OA severity was expressed in terms of the difference between the HJS-area of the osteoarthritic (HJS\_Area\_path) and contralateral normal (HJS\_Area\_norm) hip, according to the equation:

$$\text{OA severity} (\%) = \frac{\text{HJS Area}_{\text{norm}} - \text{HJS Area}_{\text{path}}}{\text{HJS Area}_{\text{norm}}} \times 100$$

(7)

Thus, OA severity takes values between 0 and 100, expressing the percentage reduction of the HJS-area due to OA. Considering that: (i) HJS-narrowing is regarded as the most reliable index for monitoring OA-progression [8], (ii) the unaffected hip in patients with unilateral OA can be used as reference for normal HJS-area [11] and (iii) in normal individuals, right and left hips do not differ significantly [36], the HJS-area difference may be employed for evaluating the cartilage loss due to OA, and thus assessing OA severity for unilateral hip OA patients. According to our clinical experience, one-fourth of patients with unilateral hip OA eventually develop OA in the contralateral hip joint.

Second, utilizing multiple linear regression analysis [37], a multivariate model of TERL features was generated for the quantitative assessment of OA using the “Matlab Statistics Toolbox”.

Following multiple trials for different combinations of textural features, the Texture Energy Run Length features TERL4 and TERL5 of the pathological HJS-ROIs were found to give best fit and they were employed as the independent (predictor) variables of the model, while the OA severity was used as the dependent variable. The regression surface describing the model was expressed by the equation:

$$\text{OA severity} (\%) = 100.880 \times \text{TERL5} - 0.028 \times \text{TERL4} + 20.555$$

(8)

where the textural features TERL4 and TERL5 are shown in Eq. (3). The quality of the regression OA severity prediction in out-of-sample data was tested according to a “Leave-3-Out” model. In particular, data for three patients were reserved for verification purposes and evaluated by the model developed using the remaining 15 patients.

2.7. Statistical analysis

The existence of statistically significant differences between normal and osteoarthritic hips for the generated textural features values was investigated employing students paired t-test. Intra-observer reproducibility concerning the manual delineation of the HJS-ROI and the subsequent
textural features generation was assessed by means of the coefficient of variation (CV) [38]. Low coefficient values correspond to high degree of reproducibility. HJS-ROIs were delineated in all radiographs twice by the same user with about a month’s interval between evaluations. For each of the twice delineated HJS-ROIs, the corresponding number of pixels was determined and the CV was calculated for these values. Student’s paired t-test was used in order to investigate whether textural features extracted from the two measurements differed significantly. Significance level was set at \( p < 0.05 \). Multiple linear regression analysis [37] was utilized for the generation of the multivariate model of TERL features for the quantitative assessment of OA. Differences between the OA severity values predicted by the regression model, developed using 18 patients, and the OA severity values estimated by the model, formed by using 15 patients (“Leave-3-Out” model), were examined by means of the student’s paired t-test. All statistical processing was performed utilizing the “Matlab Statistics Toolbox”.

### 3. Results

Statistically significant differences were derived between osteoarthritic and contralateral-normal hips for textural feature values extracted from the segmented HJS-ROIs. The highest statistically significant differences were found for all the Laws’ and TERL features (\( p < 0.001 \)).

Regarding the reproducibility of HJS-ROI determination, the CV was found equal to 3.3% indicating the reliability of the delineation process. In addition, the textural feature values that were extracted from the twice-determined HJS-ROIs, were found not to differ significantly (\( p > 0.05 \)).

### Table 1

<table>
<thead>
<tr>
<th>Textural feature</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Overall accuracy (%)</th>
<th>Feature(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run-length features</td>
<td>88.9</td>
<td>88.9</td>
<td>88.9</td>
<td>GLNU(^a), RPERC(^b)</td>
</tr>
<tr>
<td>Laws’ Texture Energy Measures(^c)</td>
<td>94.4</td>
<td>100</td>
<td>97.2</td>
<td>W5E5, R5S5, L5L5</td>
</tr>
<tr>
<td>Texture Energy Run Length features</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>TERL1(^d), TERL2, TERL4</td>
</tr>
</tbody>
</table>

\(^a\) GLNU, gray level non-uniformity.  
\(^b\) RPERC, run percentage.  
\(^c\) Using skewness statistic.  
\(^d\) Texture Energy Run Length feature.

Fig. 4. Variation of the PNN classification accuracy with the number of textural features employed, regarding the discrimination between normal and osteoarthritic hips, for the three textural feature sets studied.

Table 1 provides the highest classification accuracies between physiological and osteoarthritic hips achieved by the PNN classifier at the first level of the hierarchical tree. Similarly, Table 2 gives the discrimination precisions of the PNN classifier between mild/moderate and severe osteoarthritic hips at the second level of the hierarchical tree. Classification performance was evaluated in terms of sensitivity, specificity and overall accuracy, employing the LOO method, and for different sets of textural features. In the last columns of Tables 1 and 2, best feature combinations for achieving highest classification accuracies per feature sets are provided. The optimal value for the sigma (\( \sigma \)) parameter of the PNN classifier was determined after multiple trials to be equal to 0.24 (\( \sigma = 0.24 \)) for both levels of the hierarchical tree structure.

Fig. 4 shows the variation of the PNN classification accuracies, for level 1 of the hierarchical tree structure with increasing number of textural feature combinations and for various features sets. Fig. 5 shows the three-dimensional scatter diagram of the Texture Energy Run Length features combination (TERL1, TERL2 and TERL4) that achieved 100% classification accuracy between normal and osteoarthritic hips. Fig. 6

### Table 2

<table>
<thead>
<tr>
<th>Textural feature</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Overall accuracy (%)</th>
<th>Feature(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run-length features</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>SRE(^a), RPERC(^b)</td>
</tr>
<tr>
<td>Laws’ Texture Energy Measures(^c)</td>
<td>88.9</td>
<td>100</td>
<td>94.4</td>
<td>ESL5, SSL5, WSL5, RSR5</td>
</tr>
<tr>
<td>Texture Energy Run Length features</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>TERL3(^d), TERL4, TERL5</td>
</tr>
</tbody>
</table>

\(^a\) Short runs emphasis.  
\(^b\) Run percentage.  
\(^c\) Using skewness statistic.  
\(^d\) Texture Energy Run Length feature.
I. Boniatis et al. / Medical Engineering & Physics 29 (2007) 227–237

Fig. 5. Three-dimensional scatter diagram and decision surface of the PNN classifier for discriminating physiological (○) from osteoarthritic (□) hips employing the optimum feature combination (TERL1, TERL2 and TERL4).

Fig. 6. Variation of the PNN classification accuracy with the number of textural features employed, regarding the discrimination between hips of mild/moderate and of severe osteoarthritis, for the three textural feature sets studied.

shows the variation of the PNN classification accuracies for level 2 of the hierarchical tree structure with increasing number of textural feature combinations and for various features sets. Fig. 7 shows the three-dimensional scatter diagram of the

Table 3
Truth table demonstrating classification results for the hierarchical tree structure employing the run-length features

<table>
<thead>
<tr>
<th>Osteoarthritis severity category</th>
<th>Normal</th>
<th>Mild/moderate</th>
<th>Severe</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>88.9</td>
</tr>
<tr>
<td>Mild/moderate</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>44.4</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>66.7</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td></td>
<td>72.2</td>
</tr>
</tbody>
</table>

Table 4
Truth table demonstrating classification results for the hierarchical tree structure employing the Laws’ Texture Energy Measure features

<table>
<thead>
<tr>
<th>Osteoarthritis severity category</th>
<th>Normal</th>
<th>Mild/moderate</th>
<th>Severe</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Mild/moderate</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>88.9</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>88.9</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td></td>
<td>94.4</td>
</tr>
</tbody>
</table>

Table 5
Truth table demonstrating classification results for the hierarchical tree structure employing the Texture Energy Run Length features

<table>
<thead>
<tr>
<th>Osteoarthritis severity category</th>
<th>Normal</th>
<th>Mild/moderate</th>
<th>Severe</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Mild/moderate</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Texture Energy Run Length features combination (TERL3, TERL4 and TERL5) that achieved 100% classification accuracy between hips of mild/moderate and severe OA, respectively.

Tables 3–5 represent the truth tables for the hierarchical tree structure and for the best feature combinations of the run-length, Laws’ Texture Energy Measures and TERL feature sets, respectively.

Table 6 displays the measured OA severity values, as expressed by HJS-narrowing, and the corresponding predicted by the regression model (see Eq. (8)) OA severity values. In the last column of Table 6, OA severity values predicted by the “Leave-3-Out” model are provided. The regression surface describing the model was expressed by Eq. (8) and is shown in Fig. 8. Goodness of fit was found to be adequately satisfactory ($r=0.85$, $p<0.001$).

4. Discussion

This work proposes a texture based pattern recognition approach for classifying radiographic hip-images into OA severity categories and quantifying the severity of the disease.

Taking under consideration that the radiographic HJS comprises normal and osteoarthritic anatomical structures, the textural properties of the specific anatomical region are
Table 6
Measured and predicted values of OA severity

<table>
<thead>
<tr>
<th>Patient</th>
<th>Measured OA severity value (%)</th>
<th>Textural feature TERL4</th>
<th>Textural feature TERL5</th>
<th>Predicted OA severity value (%)</th>
<th>Predicted OA severity value by “Leave-3-Out” model (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77.2</td>
<td>714.0</td>
<td>0.8</td>
<td>81.2</td>
<td>88.2</td>
</tr>
<tr>
<td>2</td>
<td>60.0</td>
<td>484.7</td>
<td>0.6</td>
<td>67.5</td>
<td>70.1</td>
</tr>
<tr>
<td>3</td>
<td>85.5</td>
<td>284.3</td>
<td>0.8</td>
<td>93.3</td>
<td>95.3</td>
</tr>
<tr>
<td>4</td>
<td>80.4</td>
<td>598.9</td>
<td>0.8</td>
<td>84.4</td>
<td>85.6</td>
</tr>
<tr>
<td>5</td>
<td>63.7</td>
<td>665.9</td>
<td>0.6</td>
<td>62.4</td>
<td>65.5</td>
</tr>
<tr>
<td>6</td>
<td>93.0</td>
<td>168.8</td>
<td>0.7</td>
<td>86.4</td>
<td>82.5</td>
</tr>
<tr>
<td>7</td>
<td>79.1</td>
<td>522.5</td>
<td>0.7</td>
<td>76.5</td>
<td>76.6</td>
</tr>
<tr>
<td>8</td>
<td>77.4</td>
<td>478.4</td>
<td>0.7</td>
<td>77.7</td>
<td>73.4</td>
</tr>
<tr>
<td>9</td>
<td>78.4</td>
<td>730.0</td>
<td>0.7</td>
<td>70.7</td>
<td>64.1</td>
</tr>
<tr>
<td>10</td>
<td>67.4</td>
<td>599.9</td>
<td>0.6</td>
<td>64.2</td>
<td>67.8</td>
</tr>
<tr>
<td>11</td>
<td>25.9</td>
<td>599.2</td>
<td>0.5</td>
<td>54.2</td>
<td>54.4</td>
</tr>
<tr>
<td>12</td>
<td>75.0</td>
<td>388.2</td>
<td>0.5</td>
<td>60.1</td>
<td>60.7</td>
</tr>
<tr>
<td>13</td>
<td>72.2</td>
<td>668.3</td>
<td>0.6</td>
<td>62.3</td>
<td>63.9</td>
</tr>
<tr>
<td>14</td>
<td>53.9</td>
<td>782.2</td>
<td>0.5</td>
<td>49.0</td>
<td>43.3</td>
</tr>
<tr>
<td>15</td>
<td>58.7</td>
<td>762.2</td>
<td>0.5</td>
<td>49.6</td>
<td>44.0</td>
</tr>
<tr>
<td>16</td>
<td>50.5</td>
<td>707.2</td>
<td>0.5</td>
<td>51.1</td>
<td>56.1</td>
</tr>
<tr>
<td>17</td>
<td>47.2</td>
<td>470.5</td>
<td>0.4</td>
<td>47.7</td>
<td>54.0</td>
</tr>
<tr>
<td>18</td>
<td>33.6</td>
<td>722.1</td>
<td>0.4</td>
<td>40.6</td>
<td>45.8</td>
</tr>
</tbody>
</table>

* Equation of multiple linear regression model: OA severity (%) = 100.880 × TERL5 − 0.028 × TERL4 + 20.555.

expected to be altered due to osteoarthritis. These alterations were reflected by the existence of statistically significant differences in the textural feature values between the osteoarthritic and contralateral-normal hips. Highest statistically significant differences were observed for the Laws’ and the TERL feature values ($p < 0.001$). It has to be marked that the results of texture analysis of radiographic HJS could be affected by the presence of osteopetrosis, osteoporosis, and alterations in the bone mineral density of anatomical structures contained within the specific region. However, in our study only osteoarthritic patients were included, so our results refer only to HJS-alterations related to OA.

At the first level of the hierarchical tree structure, the utilization of the run-length features resulted in an overall classification accuracy of 88.9%, regarding the discrimination between normal and osteoarthritic hips (see Table 1). Two osteoarthritic hips were misclassified, giving a sensitivity accuracy of 88.9%, and two normal hips were classified as osteoarthritic, resulting in a specificity accuracy of 88.9%. Employing the Laws’ skewness textural features, overall classification accuracy was 97.2%, classifying correctly all normal hips (100% specificity), while only one osteoarthritic hip was misclassified (94.4% sensitivity). As it may also be observed from Table 1, the type of textural features employed plays an important role in the classification accuracy achieved by the classifier. Since Laws’ textural features improved the classification performance as compared to the run-length features, a combination was attempted in the introduction of a new set of textural features, as shown in Eq. (3), which classified correctly all hips. As presented in Fig. 4 and Table 1, the highest possible classification performance (100%) was achieved by employing the feature combination (TERL1, TERL2 and TERL4). The specific feature combination was found to be optimum, giving the highest classification accuracy with the minimum number of features. Fig. 5 shows the scattered diagram of the best TERL feature combination and the non-linear decision surface of the PNN classifier separating the normal from the osteoarthritic classes.

Regarding the characterization of a hip as normal or osteoarthritic, the present study introduces an advanced classification scheme, employing the PNN classifier and textural features extracted from the region of the radiographic HJS. At the second level of the hierarchical tree structure, the highest possible accuracy (100%) in discriminating hips of mild/moderate OA from hips of severe OA was achieved by employing the optimum combination TERL3, TERL4 and TERL5 (see Table 2 and Fig. 6) of the proposed TERL features. Fig. 7 shows the scattered diagram of the best TERL
feature combination and the non-linear decision surface of the PNN classifier separating the mild/moderate OA from the severe OA classes.

In contrast, the Laws’ features achieved the second best score (94.4%) classifying properly all hips of mild/moderate OA (100% specificity), while only one hip of severe OA was misclassified (88.9% sensitivity). The run-length features performed poorly, since overall classification accuracy was 66.7%. Six hips of mild/moderate OA were classified properly, resulting in a specificity accuracy of 66.7%, while three hips of severe OA were misclassified, giving a sensitivity accuracy of 66.7%.

The overall classification accuracies of the hierarchical tree achieved employing: (i) the run-length, (ii) the Laws’ Texture Energy Measures and (iii) the TERL feature sets were 72.2, 94.4 and 100%, respectively (see Tables 3–5).

Finally, in order to assess quantitatively the severity of hip OA and, perhaps, to potentially monitor the progression of the disease in the same patient, a regression mathematical model was derived employing multiple linear regression analysis [37]. The resulting regression Eq. (8) comprised TERL4 and TERL5 textural features, which evaluate textural properties (variability of organized structure and texture homogeneity, respectively [31,32]) of the radiographic HJS of osteoarthritic hips. The graphical representation of the relationship between the variables involved in Eq. (8) is shown in Fig. 8. As it can be observed from Fig. 8 and Table 6, plotted data followed a well-behaved regression plane. The fact that no significant differences were found between the values obtained by the model of Eq. (8) and those obtained by the “Leave-3-Out” model ($p > 0.05$), it may be suggestive that the proposed prediction model may be trusted for out-of-sample data, as is the case in a practical application. Taking into account that the model utilizes textural features extracted only from osteoarthritic HJSs, it may be concluded that the proposed method can be of value in the monitoring of the disease progression. However, considering that our approach was developed utilizing a small sized data set, further validation of the proposed model(s) is required employing a larger patient data set in the future.

5. Conclusion

In conclusion, the proposed classification system achieved high accuracy in the characterization of hips as normal or osteoarthritic and in the discrimination among various grades of hip OA severity so that it may be adopted as a decision support tool. Additionally, a regression model was derived that quantifies OA severity reliably and could be employed in the monitoring of OA progression. Finally, the classification system and regression model were both designed employing novel textural features, which reflected properties of organized structure and radiographic texture in the HJS, and demonstrated texture alterations occurring in the hip joint due to osteoarthritis.

Acknowledgements

The first author was supported by a grant from the State Scholarship Foundation (SSF), Greece. The authors thank the staff of the Departments of Orthopaedics and Radiology for their contribution to this work.

References


