Med Bio Eng Comput (2006) 44:793–803
DOI 10.1007/s11517-006-0096-3

ORIGINAL ARTICLE

Osteoarthritis severity of the hip by computer-aided grading of radiographic images

I. Boniatis · L. Costaridou · D. Cavouras · I. Kalatzis · E. Panagiotopoulos · G. Panayiotakis

Received: 19 December 2005 / Accepted: 20 July 2006 / Published online: 15 August 2006

Abstract A computer-aided classification system was developed for the assessment of the severity of hip osteoarthritis (OA) . Sixty-four radiographic images of normal and osteoarthritic hips were digitized and enhanced. Employing the Kellgren and Lawrence scale, the hips were grouped by three experienced orthopaedists into three OA-severity categories: Normal, Mild/Moderate and Severe. Utilizing custom-developed software, 64 ROIs corresponding to the radiographic Hip Joint Spaces were manually segmented and novel textural features were generated. These features were used in the design of a two-level classification scheme for characterizing hips as normal or osteoarthritic (1st level) and as of Mild/Moderate or Severe OA (2nd level). At each classification level, an ensemble of three classifiers was implemented. The proposed classification scheme discriminated correctly all normal hips from osteoarthritic hips (100% accuracy), while the discrimination accuracy between Mild/Moderate and Severe osteoarthritic hips was 95.7%. The proposed system could be used as a diagnosis decision-supporting tool.

I. Boniatis · L. Costaridou · G. Panayiotakis (✉)
Department of Medical Physics, School of Medicine, University of Patras, 265 00 Patras, Greece
e-mail: panayiot@upatras.gr

D. Cavouras · I. Kalatzis
Department of Medical Instrumentation Technology, Technological Educational Institute of Athens, 122 10 Athens, Greece

E. Panagiotopoulos
Department of Orthopaedics, School of Medicine, University of Patras, 265 00 Patras, Greece

Keywords Hip · Osteoarthritis · Radiography · Texture · Pattern recognition

Abbreviations

OA	Osteoarthritis
HJS	Hip joint space
KL	Kellgren and Lawrence
ROI	Region of interest
CLAHE	Contrast limited adaptive histogram equalization
GEN_Image	Gabor energy image
GLRLM	Grey level run length matrix
GEMRL	Gabor energy measure run length
PNN	Probabilistic neural network
k-NN	*k*-Nearest–neighbour
MV	Majority vote
CV	Coefficient of variation

1 Introduction

Osteoarthritis (OA) is considered to be the most common cause of human disability in industrialized countries [1]. The condition is characterized by progressive disintegration and loss of the articular cartilage, accompanied by alterations in the subchondral joint tissues [22].

Although Magnetic Resonance Imaging is regarded as the most sensitive tool for assessing articular cartilage [26], plain film radiography is considered as an adequate resource for the initial clinical evaluation of the disease [23]. The dominant radiographic features of

hip OA, reflecting the degeneration of the articular cartilage and alteration in the subchondral bone tissue, comprise Hip Joint Space (HJS) narrowing, subchondral cysts development, osteophyte formation and subchondral sclerosis [3].

The radiographic assessment of OA concerns the confirmation of the existence of the disease, the assessment of the severity as well as the monitoring of the progression of joint structural alterations. For the assessment of the severity of hip OA, several radiographic scoring systems have been proposed. These scoring systems are mainly based on the subjective assignment of a severity grade to the studied hip joint, while the definitions of severity grades are based on aspects of joint structural alterations visualized on plain radiographs [31]. Among the proposed radiographic scoring systems, the Kellgren and Lawrence (KL) grading scale [19] is considered as the gold standard for epidemiological studies [30]. The KL scale defines five categories of OA severity ranging between 0 and 4, with grades ≥ 2 corresponding to osteoarthritic pathology. Hips with KL = 0 or 1 are characterized as normal or doubtful for OA, respectively, while the grades 2, 3 and 4 are assigned to hips of mild, moderate and severe OA, respectively [19].

Monitoring of hip OA progression refers to the measurement of joint structural–anatomic changes with time, which are associated with the degenerative action of the disease. Specifically, progression mostly refers to the quantitative assessment of articular cartilage loss and HJS narrowing is considered the most reliable index for evaluating the progression [3]. Quantification of HJS narrowing concerns manual or computerized measurements of radiographic HJS parameters, such as HJS width and/or HJS area on serial radiographs [11, 15].

Texture as a regional descriptor of a digital image is related to the distribution and spatial interrelationships of pixel intensities corresponding to the region. Computer-based texture analysis of digital images concerns the utilization of algorithms capable of quantifying the textural properties of an image [10]. In the field of medical imaging, texture analysis can be used for the extraction of diagnostically meaningful information by means of textural features, which may not be easily perceivable [33]. In addition, the extracted information can be used for distinguishing between tissue structures, which are represented by textures in medical images [4, 34].

Within the framework of the radiographic investigation of hip OA, a computer-based method has recently been proposed by our group for the quantitative assessment of the severity of hip OA and the monitoring of the progression of the disease . The method refers to the utilization of first and higher order statistical textural features, extracted from the region of the radiographic HJS, and the introduction of textural-features thresholds for grading and quantifying severity of hip OA [7].

In this study, a more robust pattern recognition approach, based on classification algorithms and textural features calculated from Gabor filters, is proposed for the assessment of the severity of hip OA . To the best of our knowledge, a texture-based pattern recognition approach has not been reported for the specific purpose. The proposed method concerns: (a) the calculation of novel textural features extracted from the region of radiographic HJS, (b) the implementation of classification algorithms (classifiers) for characterizing hips as Normal or Mild/Moderate or Severe OA, and (c) the combination of different classifiers (ensembles) in order to improve the diagnostic accuracy of the method.

2 Materials and methods

2.1 Patients and radiographs

The sample of the study comprised 64 radiographic hip joint images (18 normal and 46 osteoarthritic), corresponding to 32 patients with verified unilateral or bilateral hip OA. Patients' ages ranged between 49 and 83 years, with a mean age of 66.7 years. Eighteen patients were diagnosed as unilateral OA while 14 as bilateral OA. The American College of Rheumatology criteria were used for OA diagnosis. In particular, and within the context of combined clinical and radiographic criteria for OA diagnosis, the characterization of a hip as osteoarthritic was based on the restricted mobility of the joint as well as on the presence of pain (associated to hip joint use) with the combined radiographic evidence of osteophytes and/or joint space narrowing [2]. This combined evaluation of OA was necessary, since disparity may exist between a radiographic score and a clinical finding of OA.

All radiographs were taken employing the same radiographic protocol, which comprised use of the same X-ray unit (Siemens, Polydoros 50, Erlangen, Germany), alignment of the X-ray beam 2 cm above the pubic symphysis, a film-focus distance of 100 cm, tube voltage between 70 and 80 kVp and use of a fast screen and film cassette (30×40 cm^2). Pelvic radiographs were digitized employing a laser digitizer, suitable for medical applications [21] (Lumisys, Sunnyvale,

CA, USA). Digitizer performance was evaluated employing a quality control protocol [13]. It is obvious that in the case of digital radiography systems this intermediate scanning step is avoided.

The severity of OA was assessed radiographically by three experienced orthopaedists, employing the KL grading scale [19]. Each of the orthopaedists graded OA by assigning a KL grade to the examined hips, while radiographs were assessed at two different time points with about a month's interval between the evaluations. In order to establish a golden-standard, and thus to design our system, only those exams of common consent were retained for the purposes of the present study. In this context, from the 74 radiographic images of the original sample, 64 were retained for the remaining part of the study. Based on the KL scale, the hips of the sample were grouped into three major OA-severity categories*: Normal/Doubtful* (KL = 0, 1), *Mild/Moderate* (KL = 2, 3) and *Severe* (KL = 4). Accordingly, 18 hips were assigned to *Normal/Doubtful category*, 16 to *Mild/Moderate* and 30 to *Severe.*

2.2 Radiograph enhancement and determination of radiographic hip joint spaces

On each pelvic radiograph, two Regions Of Interest (ROIs) corresponding to a patient's both HJSs were determined, employing custom-developed software [27, 28]. Specifically, an algorithm realizing the Contrast-Limited Adaptive Histogram Equalization (CLAHE) method [24] was employed in MATLAB (The MathWorks Inc., Natick, MA, USA) to enhance the contrast of the digitized radiographs, and thus to emphasize the articular margins of the hip joint. As it can be seen in Fig. 1, an acute angle encompassing the weight-bearing portion of the hip joint provided the medial and lateral limits of the HJS-ROI [11]. These limits as well as the articular margins of the joint (edge of the femoral head, inferior margin of the acetabulum) were individually determined by each orthopaedists. This ROI (Fig. 2) was used for the extraction of textural features. Only ROIs segmented by the orthopaedist with the highest reproducibility were used.

Following the above-mentioned procedure for each patient of the sample, 64 HJS-ROIs were obtained in total. From these, 18 HJS-ROIs corresponded to the contralateral normal hips of 18 unilateral OA patients. These normal HJS-ROIs were used as the control sample of the study (Normal/Doubtful category). The remaining 46 ROIs corresponded to the 18 and the 28 pathological hips of the unilateral and bilateral OA patients, respectively. These ROIs were distributed to the Mild/Moderate and the Severe categories.

2.3 Textural features generation

Employing custom-developed algorithms in MATLAB, textural features were generated from the segmented HJS-ROIs employing the multi-channel two-dimensional (2-D) Gabor filtering approach [12]. Additionally, novel textural features, aiming to capture organized structures within the region of radiographic HJS, were introduced in the present study. These features, labelled as ''Gabor Energy Measure Run Length'', were generated by the combination of the Grey Level Run Length Matrix (GLRLM) method [14] and the 2-D Gabor filtering approach [12].

2.3.1 The multi-channel two-dimensional Gabor filtering approach

According to the multi-channel Gabor filtering approach, textural features (Gabor features) can be obtained by filtering an input image with a set of 2-D Gabor filters. A 2-D Gabor filter can be considered as a sinusoidal plane wave of certain spatial frequency and orientation that is modulated by a 2-D Gaussian envelope. Gabor filters provide the advantage that can be simultaneously be optimally localized in both the spatial and spatial frequency domains [12].

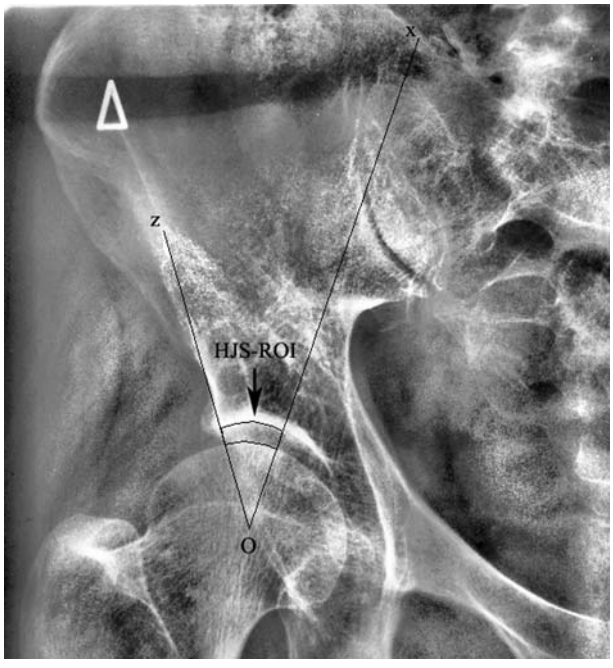

Fig. 1 Determination of the hip joint space ROI. *xOz* acute angle defined by patient's standard anatomical landmarks encompassing the examined ROIs. *x* highest point of the homolateral sacral wing, *z* lateral rim of the acetabulum, *O* centre of the femoral head

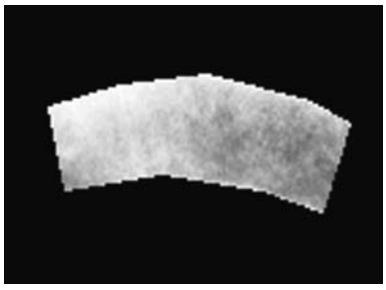

Fig. 2 Example of segmented hip joint space ROI corresponding to Fig. 1

For the needs of the present study, 5×5 Gabor filters were used while four (4) filtered images, corresponding to filter orientations of 0, 45, 90 and 135°, were generated for each HJS-ROI image (see Fig. 3). In these images, characterized as ''Gabor Energy Images (GEN_Images),'' the pixel intensities represent the values of a measurement, labelled as Gabor Energy [16]. From each one of the GEN_Images, six first-order statistics (mean θ, variance θ, skewness θ, kurtosis θ, range θ and standard deviation θ, where, θ: 0, 45, 90 and 135° denotes the orientation of the Gabor filter) were computed providing 24 Gabor first order statistics textural features. Details concerning the computation of the Gabor features are given in the Appendix.

2.3.2 Gabor energy measure run length features

The GLRLM method [14] and the 2-D Gabor filtering approach [12] were suitably combined to provide novel textural features in an effort to improve the Gabor features discriminatory capacity.

The suggested approach comprised the following steps: (a) The HJS-ROI image was convolved with the 2-D Gabor filter. (b) The GEN_Image of the HJS-ROI was formed by replacing every pixel in the filtered image with a Gabor Energy measure associated to the pixel [16]. (c) As described by Galloway [14], the GLRLM along the angular direction d of the GEN_Image was computed (d = 0, 45, 90 and 135°, while d is defined in a similar way as θ for the case of Gabor filters) and five textural features, labelled as ''Gabor Energy Measure Run Length (GEMRL_θ)'' features, were extracted (θ: 0, 45, 90 and 135° denotes the orientation of the Gabor filter that is implemented on the HJS-ROI image). Considering that five GEMRL features were computed from each one of the four GEN_Images, 20 novel textural features were produced for each HJS-ROI. Details concerning the computation of GEMRL features can be found in the Appendix.

2.4 Feature normalization and reduction

All textural features employed in the present study (24 Gabor first order statistics and 20 GEMRL) were normalized to zero mean and unit standard deviation according to Eq. 1:

$$\mathrm{ft}_{k\_\mathrm{norm}} = \frac{\mathrm{ft}_k - \mu}{\sigma}, \tag{1}$$

where $\mathrm{ft}_{k\_\mathrm{norm}}$ is the normalized value of the ft_k textural feature, while μ and σ are the mean value and standard deviation, respectively, of feature ft_k over all HJS-ROIs. Each time that an isolated case is fed into the system, the same normalization procedure is followed by calculating the μ and σ values over all HJS-ROIs but the isolated one. The latter is then normalized employing the μ and σ values calculated by the rest of the patterns.

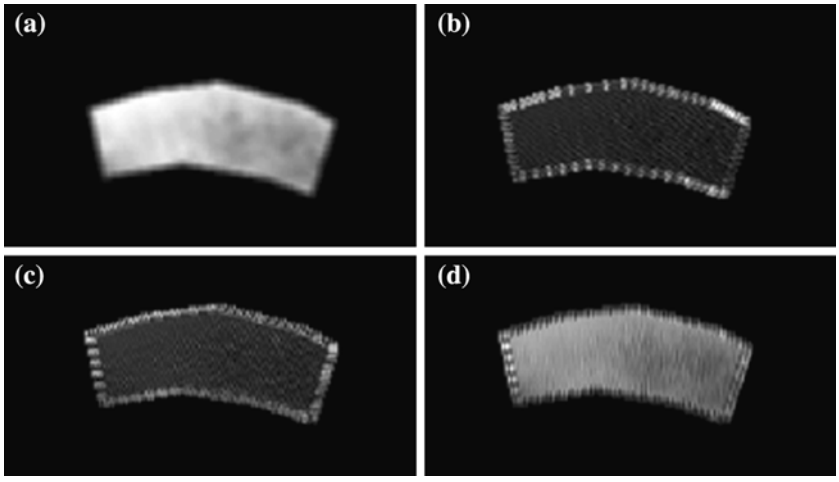

Fig. 3 Gabor Energy Images (*GEN_Images*) of the segmented hip joint space ROI, generated as the result of filtering Fig. 2 with Gabor filters of: (**a**) 0° orientation (GEN_Image_0°), (**b**) 45° orientation (GEN_Image_45°), (**c**) 90° orientation (GEN_Image_90°) and (**d**) 135° orientation (GEN_Image_135°)

The discriminatory ability of each one of the textural features was tested employing the Student's *t*-test. Only the best discriminating features ($p < 0.001$) were selected and were employed in the design of the classification scheme [32]. For each classification task, the optimum combination of the best discriminating features (i.e., the combination providing the highest classification accuracy employing the minimum number of features) was determined according to the exhaustive search procedure [32]. More specifically, every possible feature combination (i.e., combinations of 2–5 features) was used for the design of the classifier, while the performance of the classifier was tested each time. Finally, the feature combination that demonstrated the highest classification accuracy with the smallest number of features was selected as the optimum one.

2.5 Classification of hips into osteoarthritis-severity categories

A hierarchical decision tree was developed for the assignment of hips into three OA-severity categories labelled as Normal, Mild/Moderate and Severe. Figure 4 shows the structure of the classification approach employed in the present study. The classification system 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 as osteoarthritic in the first level were further classified as of Mild/Moderate or Severe OA. Each level was implemented by means of a combined classification scheme employing the following individual classifiers: Probabilistic Neural Network (PNN) classifier [29], *k*-Nearest–Neighbour (*k*-NN) classifier and Bayes classifier [32]. Accordingly, at each level of the decision tree, the textural features that were extracted from the region of radiographic HJS were used as inputs to each one of the individual classifiers, while the classifiers' outputs corresponding to the best classification performances were combined employing the Majority Vote (MV) rule [5, 18]. According to the latter, an unknown pattern is assigned to a specific class ω_k if the majority of individual classifiers, forming the combined classifier, assigns the pattern to ω_k.

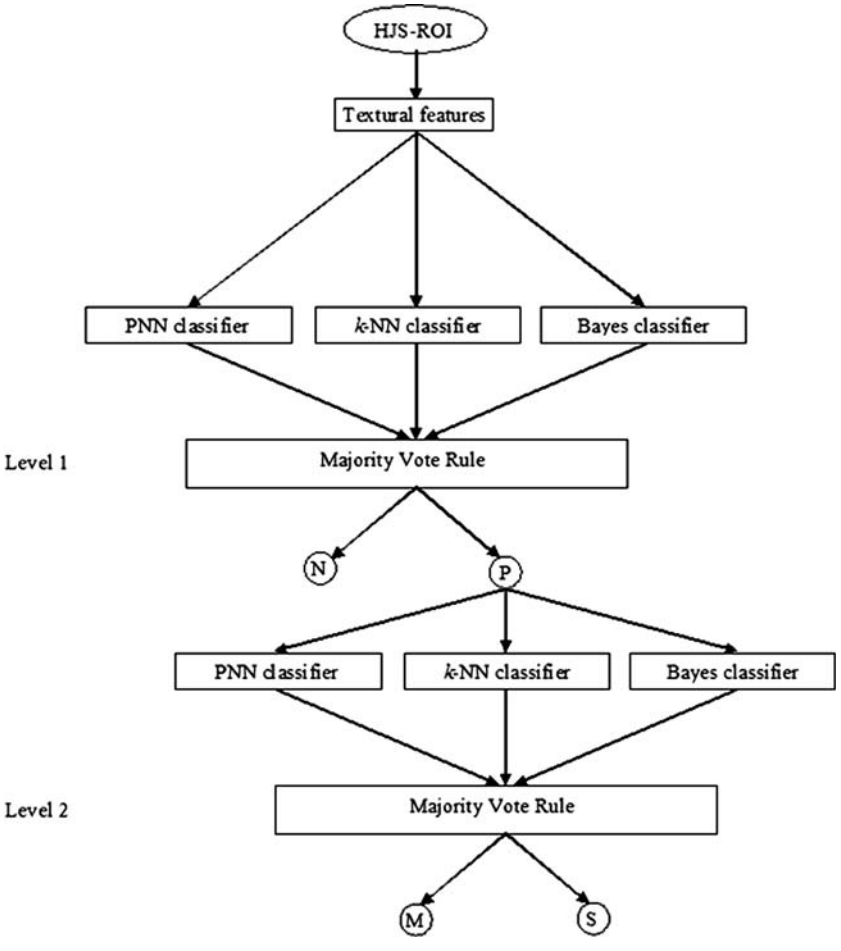

Fig. 4 The hierarchical decision tree structure for the discrimination between normal (*N*) and osteoarthritic (*P*) hips at Level 1, and between hips of Mild/Moderate (*M*) and Severe (*S*) osteoarthritis at Level 2. In both levels, the PNN, *k*-NN and Bayes classifiers were combined employing the MV rule

2.5.1 Probabilistic neural network classifier

The PNN classifier, as proposed by Specht [29], encompasses both the Bayes' classification approach and the Parzen's estimators of probability density functions. PNN's main advantages are that: (a) it is fast to train and (b) the probability density functions of the pattern vectors forming a class do not have to follow a Gaussian distribution [29]. The decision function of the PNN classifier employed in this study is described by Eq. 2:

$$d_k(\mathbf{x}) = \frac{1}{(2\cdot\pi)^{n/2}\cdot\sigma^n\cdot N_k}\sum_{i=1}^{N_k} e^{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\cdot\sigma^2}}, \quad (2)$$

where $\mathbf{x}_i$ is the ith training input pattern, $\mathbf{x}$ is the unknown pattern to be classified, N_k is the number of patterns forming the class ω_k, n is the number of textural features forming the input pattern while sigma (σ) is an adjusting parameter, taking values ranging between 0 and 1. According to Eq. 2, as the distance between $\mathbf{x}$ and $\mathbf{x}_i$ ($\|x\text{-}x_i\|$) increases, the exponential term approaches 0, indicating a small similarity between the two pattern vectors. On the other hand, as the distance between $\mathbf{x}$ and $\mathbf{x}_i$ ($\|x\text{-}x_i\|$) decreases, the exponential term approaches 1, indicating a significant similarity between the two pattern vectors. As sigma approaches 0, even small differences between $\mathbf{x}_i$ and $\mathbf{x}$ will provide a zero value for the exponential term, while larger values of sigma provide more smooth results. As it can be concluded, the selection of sigma affects the estimation error of the PNN and is determined experimentally by comparing the accuracies obtained for different values of the parameter. The unknown pattern $\mathbf{x}$ was classified to the class with the highest value of decision function d_k ($\mathbf{x}$) [29, 32].

2.5.2 Bayes classifier

Bayes classification is based on decision functions as described by Eq. 3:

$$d_k(\mathbf{x}) = p(\mathbf{x}|\omega_k)\cdot P(\omega_k), \quad k = 1,2,...,N, \quad (3)$$

where $p(\mathbf{x}|\omega_k)$ is the probability density function of the pattern vectors of class ω_k, while $P(\omega_k)$ represents the probability concerning the occurrence of class ω_k. Considering an unknown pattern, the latter is assigned to the class whose decision function provides the highest value. Assuming equal probabilities $P(\omega_k)$ and Gaussian probability density functions $p(\mathbf{x}|\omega_k)$, the decision function of the classifier may be represented by Eq. 4:

$$d_k(\mathbf{x}) = \ln P(\omega_k) - \frac{1}{2}\cdot\ln |\mathbf{C}_k| - \frac{1}{2}\left[(\mathbf{x}-\mathbf{m}_k)^T\cdot\mathbf{C}_k^{-1}\cdot(\mathbf{x}-\mathbf{m}_k)\right], \quad k = 1,2,...,N, \quad (4)$$

where $\mathbf{C}_k$ and $\mathbf{m}_k$ represent the covariance matrix and the mean vector of class ω_k, while T indicates transposition [32].

2.5.3 k-Nearest–neighbour classifier

The k-Nearest–Neighbour (k-NN) classification algorithm is one of the most successful routines for performing general, non-parametric classification, since no prior knowledge of patterns distributions is required.

According to the k-NN classification approach, an unknown pattern is assigned to the class where the majority of its k-Nearest–Neighbours belong. The term ''nearest'' is related to the lowest value of a predefined distance function. For a two-classes classification task, as the number of neighbours increases ($k \to \infty$) the performance of the classifier tends to the optimal one. For the needs of this study, the Euclidean distance function was used [32]. Referring to the n-dimensional feature space, the Euclidean distance between training and the unknown pattern is defined according to Eq. 5:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\left[(x_1-y_1)^2 + (x_2-y_2)^2 + \cdots + (x_n-y_n)^2\right]}, \quad (5)$$

where, $d(\mathbf{x},\mathbf{y})$ is the Euclidian distance, $\mathbf{x} = [x_1\ x_2\ ...\ x_n]$ represents the pattern vector to be classified, while $\mathbf{y} = [y_1\ y_2\ ...\ y_n]$ denotes the training pattern.

Each classifier was validated for both levels employing the Leave-One-Out 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. In this way, the classifier is evaluated by data that are not involved in its design. The specific validation method was used, since an independent data set that could be used for the verification of our results was not accessible. The performance of the combined classification scheme was evaluated in terms of sensitivity, specificity and overall accuracy [32].

2.6 Statistical analysis

Student's t-test was used in order to investigate the existence of statistically significant differences between osteoarthritic and normal hips for textural feature

values. In addition, the Student's *t*-test was employed for the evaluation of the discriminatory ability of each one of the textural features. In particular, the textural features that provided a p value lower than 0.001 ($p < 0.001$) were considered as the best discriminating ones, and thus were selected and further employed in the design of the classification scheme [32]. The specific significance level was considered as a proper one, even after the Bonferroni correction of the p value [9]. The specific correction was considered as necessary, since multiple significance tests were performed for the various textural features. Normality of distributions was verified employing the Lilliefors test [20]. Intra-observer and inter-observer reproducibility of HJS-ROI outlining were both evaluated by means of the Coefficient of Variation (CV) [35]. Accordingly, HJS-ROIs were delineated in all radiographs twice by each one of the experienced orthopaedists with about a month's interval between evaluations. In order to investigate whether textural features extracted from the two measurements differed significantly, the Student's paired *t*-test was used. The HJS-ROIs that were segmented with the highest degree of reproducibility were used for subsequent texture analysis.

3 Results

The distributions of all the textural feature values were found to be Gaussian according to the Lilliefors test [20]. Statistically significant differences ($p < 0.001$) were derived between normal and osteoarthritic hips for the various textural feature sets employed in the study. The determination of the HJS-ROIs was found to be reproducible, since the CV values regarding the intra-observer and the inter-observer reproducibility were, on an average, 3.4 and 4.2%, respectively. In addition, the textural feature values, which were extracted from the HJS-ROIs that were determined with the highest degree of reproducibility, were found not to differ significantly ($p > 0.05$).

Table 1 summarizes the best classification scores achieved by the individual classifiers regarding the discrimination between normal and osteoarthritic hips. The classification accuracies displayed correspond to the two textural feature sets used. Regarding the utilization of the Gabor first order statistics features, the PNN classifier accomplished a classification accuracy of 96.9% utilizing the feature combination comprising the features (mean_45°, standard deviation_0°, standard deviation_135°, kurtosis_135° and range_90°). For the specific case, the sigma (σ) value was determined, after multiple trials, to be equal to 0.3. The same classification score was achieved by the *k*-NN classifier for the feature combination (standard deviation_90°, skewness_ 135°, kurtosis_135°, range_90°) and 5 Nearest–Neighbours ($k = 5$). Finally, the classification performance of the Bayes classifier for the best feature vector combination (standard deviation_135°, variance_0°, kurtosis_135°) was 95.3%. Regarding the characterization of hips as normal or osteoarthritic employing the proposed GEMRL features, the PNN classifier characterized correctly 63 of the 64 hips (98.4% overall accuracy) utilizing the feature combination (GEMRL3_45°, GEMRL4_45°, GEMRL2_90° and GEMRL3_135°) for a sigma value of 0.3 ($\sigma = 0.3$). The same classification performance was accomplished by the *k*-NN classifier for the feature combination (GEMRL3_0°, GEMRL1_90°, GEMRL3_90°, GEMRL1_135° and GEMRL5_135°) and $k = 5$. Regarding the Bayes classifier, the corresponding classification accuracy was 96.9% utilizing the feature combination (GEMRL2_0°, GEMRL1_135°, GEMRL2_135° and GEMRL3_135°). The combination of the PNN and the *k*-NN classifiers, when both were designed with the GEMRL features, achieved a lower classification score (96.9%) in comparison to the MV scheme for the three classifiers (100%).

Table 2 represents the truth table concerning the discrimination between normal and osteoarthritic hips employing the combined classification scheme and the GEMRL features. The highest possible classification accuracy (100%) was achieved, since all the normal

Table 1 Highest classification accuracies regarding the discrimination between normal and osteoarthritic hips, utilizing individual classifiers and various textural feature sets

Classifier	Gabor first order statistics (%)	GEMRL features (%)[a]
PNN	96.9	98.4
k-NN	96.9	98.4
Bayes	95.3	96.9

[a] Gabor Energy Measure Run Length

Table 2 Combined classification scheme discrimination accuracy between normal and osteoarthritic hips employing the Gabor Energy Measure Run Length (GEMRL) features

Osteoarthritis severity categories	Normal	Osteoarthritic	Accuracy (%)
Normal	64	0	100
Osteoarthritic	0	64	100
Overall accuracy			100

Table 3 Combined classification scheme discrimination accuracy between hips of Mild/Moderate and of Severe osteoarthritis employing the best features (Gabor Energy Measure Run Length features)

Osteoarthritis severity categories	Mild/Moderate	Severe	Accuracy (%)
Mild/Moderate	15	1	93.8
Severe	1	29	96.7
Overall accuracy			95.7

(100% specificity) and osteoarthritic hips (100% sensitivity) were characterized correctly.

Similarly, Table 3 corresponds to the truth table for the 2nd level of the hierarchical tree structure. Employing the best features (GEMRL features), the ensemble of three classifiers assigned 44 of the 46 osteoarthritic hips to the correct category, achieving a classification accuracy of 95.7%. Only one hip of Mild/Moderate OA was misclassified, providing a specificity accuracy of 93.8%, while 29 of the 30 hips were characterized as of Severe OA, resulting in a sensitivity accuracy of 96.7%.

Table 4 tabulates the classification performance of the individual classifiers in discriminating between hips of Mild/Moderate and of Severe OA, employing the GEMRL features. The PNN classifier characterized correctly 42 of the 46 hips (91.3% overall accuracy) utilizing the feature combination (GEMRL1_0°, GEMRL4_0°, GEMRL3_45°, GEMRL1_90° and GEMRL5_135°) for $\sigma = 0.2$. The k-NN classifier provided an overall accuracy of 89.1%, assigning to the correct categories 41 of the 46 hips for the feature combination (GEMRL4_0°, GEMRL5_0°, GEMRL2_45°, GEMRL4_90° and GEMRL3_135°) and three Nearest–Neighbours ($k = 3$). Regarding the Bayes classifier, the accomplished classification score was also 89.1% for the feature combination (GEMRL2_90°, GEMRL4_90°, GEMRL1_135° and GEMRL4_135°).

Table 4 Classification performance of individual classifiers at Level 2, regarding the discrimination between hips of Mild/Moderate and of Severe osteoarthritis employing the Gabor Energy Measure Run Length features

Classifier	Overall accuracy (%)	Specificity (%)	Sensitivity (%)
PNN	91.3	93.8	90.0
k-NN	89.1	75.0	96.7
Bayes	89.1	81.3	93.3

4 Discussion

In this study, a computer-aided classification system, based on pattern recognition methods, is proposed for characterizing hips as Normal, Mild/Moderate or Severe OA from radiographic images. In digitized radiographs, texture corresponds to the spatial organization of pixel intensities variation, which is attributed to X-ray attenuation properties of anatomical structures [6]. In this study, the analysed HJS-ROI comprises either osteoarthritic and/or normal superimposed anatomical components, since it is formed by the superposition of three-dimensional anatomical structures of the articular cartilage, the posterior acetabular wall and the iliac bone. Thus, textural features extracted from the region of radiographic HJS could provide valuable textural information regarding the condition of anatomical structures in the hip joint. In OA, the articular cartilage and the subchondral bone tissues undergo biochemical, biomechanical and structural alterations [1, 22]. Degradation of the articular cartilage in terms of fibrillation, progressive softening and disintegration results in increased loading on the subchondral bone, which causes a remodelling of the bone. On the other hand, alterations in the structure as well as in the mechanical properties of subchondral bone have been associated with cartilage degeneration in OA [8, 25]. These alterations are expected to affect the attenuation properties of hip joint tissues and, as a consequence, the radiographic HJS texture. Statistical analysis revealed the existence of statistically significant differences between normal and osteoarthritic hips for the textural feature sets used in this study, signifying the alteration of HJS texture due to OA. Within this context, the Gabor first order statistics and the proposed GEMRL textural features were employed in the design of the classification schemes.

Regarding the discrimination between normal and osteoarthritic hips (see Table 1), the utilization of GEMRL features improved the classification performance of each one of the individual classifiers in comparison to the Gabor first order statistical features. In an effort to further improve the discrimination accuracy, combined classification schemes were implemented at both levels of a hierarchical decision tree. In particular, the outputs provided by the PNN, k-NN and the Bayes classifiers (each classifier utilizing its own optimum combination of GEMRL features) were combined by the MV rule [5, 18]. As Table 2 demonstrates, the proposed classification scheme achieved the highest possible classification score by discriminating correctly all normal hips from osteoarthritic hips. In a

previous study [17], the discrimination between normal and osteoarthritic hips has been performed by the application of quantitative thresholds on manually measured HJS-width values on colon radiographs, with an error rate of 3.6% for joints with HJS-width ≤ 2.5 mm. In a previous study by our group [7], employing thresholds on the Grey Level Non-Uniformity textural feature [14], the discrimination scores between normal and OA hips and among various grades of OA severity was 90.6 and 88.9%, respectively. The present study proposes a texture-based pattern recognition system for the characterization of hip OA severity.

At the 2nd level of the hierarchical decision tree, the hips that had been characterized as osteoarthritic were further discriminated as of Mild/Moderate or of Severe OA. The ensemble of classifiers improved the discrimination accuracy between hips of Mild/ Moderate and of Severe OA as compared to individual classifiers. This can be attributed to the fact that the MV rule utilizes the regions of feature space, where each classifier achieves its best classification performance [18].

As it can be observed from Tables 2 and 3, the utilization of the GEMRL features at both levels of the classification structure resulted in relatively high classifications scores regarding the discrimination among various grades of hip OA severity. This finding may be indicative of the capacity of the proposed features to differentiate between normal and osteoarthritic hips as well as between hips of Mild/Moderate OA and of Severe OA. The discriminatory power of the GEMRL features may be justified considering that their computation is based on the combination of the GLRLM method and the Gabor filtering approach. In particular, the degenerative action of OA is expected to affect the Gabor Energy measure values in the corresponding filtered HJS-ROI images. Thus, textural features involving Gabor Energy measures may be of value in the evaluation of structural alterations of the hip joint related to OA. In particular, the GEMRL features may quantify inherent textural properties of the GEN_-image that are related to the predominance of small (short runs) or large (long runs) structures and to the variability in grey level or length of the organized linear structures. Generalizing, the results of the present study indicate that the radiographic texture of HJS provides diagnostically useful information regarding the assessment of OA severity. Thus, beyond the well-known HJS width and/or HJS area, the parameters of radiographic HJS that could be used for the evaluation of osteoarthritic alterations in the hip joint can include textural features generated from the specific region. Within this context, the proposed computer-based system can be considered as a supportive tool to the diagnostic procedure followed in clinical routine, since: (a) it is compatible with the KL grading scale, and (b) it discriminates among OA-severity categories successfully, as it is indicated by the high classification scores that were accomplished. In conditions of clinical routine, the system's computer processing time is infinitesimal. However, the only time demanding operation (<1 min) concerns mostly the manual delineation of radiographic HJS by the orthopaedist.

5 Conclusions

A computer-aided classification system was developed for assessing the severity of hip OA from radiographic images. The system relies on the Kellgren and Lawrence grading scale, which is considered as the gold standard for epidemiological studies. The suggested approach concerns the generation of textural features from the region of radiographic HJS and their utilization in the design of a hierarchical decision tree structure. Specifically, textural features employing the Gabor Energy measures in radiographic HJS images may reflect the degenerative action of OA in the hip joint tissues. Considering the relatively high classification scores achieved in this study, the utilization of textural features capable of quantifying the spatial distribution and the linear organization of Gabor Energy measures within the radiographic HJS may be valuable in the discrimination between normal and osteoarthritic hips as well as among grades of hip OA severity. The proposed system could be used as a diagnosis decision-supporting tool.

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

6 Appendix

6.1 Generation of Gabor textural features

A two-dimensional (2-D) Gabor filter $G(x,y)$ can be considered as a sinusoidal plane wave of certain spatial frequency and orientation, modulated by a 2-D Gaussian envelope [12].

For the needs of this study, four (4) filter orientations were used: $\theta° = 0°$, 45°, 90°, and 135°. For each orientation θ, a pair of filters with an anti-symmetric phase relationship was used [16]: $G_{f,\theta,0°}(x,y)$ and $G_{f,\theta,-90°}(x,y)$.

Each HJS image, corresponding to the determined Region Of Interest (ROI), was convolved with the, $G_{f,\theta,0^\circ}(x,y)$ as well as with the $G_{f,\theta,-90^\circ}(x,y)$ filter, according to Eqs. 6 and 7, respectively:

$$\mathrm{GFIM}_{f,\theta,0^\circ}(i,j)=\sum_{x=-m}^{+m}\sum_{y=-m}^{+m}G_{f,\theta,0^\circ}(x,y)I(i+x,j+y),\quad m=\frac{z-1}{2},z=5. \tag{6}$$

$$\mathrm{GFIM}_{f,\theta,-90^\circ}(i,j)=\sum_{x=-m}^{+m}\sum_{y=-m}^{+m}G_{f,\theta,-90^\circ}(x,y)I(i+x,j+y),\quad m=\frac{z-1}{2},z=5, \tag{7}$$

where $I(i,j)$ is the input HJS-ROI image, $G_{f,\theta,0^\circ}(x,y)$ and $G_{f,\theta,-90^\circ}(x,y)$ are the z × z Gabor filters, while $\mathrm{GFIM}_{f,\theta,0^\circ}(i,j)$ and $\mathrm{GFIM}_{f,\ \theta,\ -90^\circ}(i,j)$ represent the filtered images corresponding to the $G_{f,\theta,0^\circ}(x,y)$ and $G_{f,\theta,-90^\circ}(x,y)$ filters, respectively.

Based on the filtered images $\mathrm{GFIM}_{f,\theta,0^\circ}(i,j)$ and $\mathrm{GFIM}_{f,\theta,-90^\circ}(i,j)$, an image labelled as Gabor Energy Image (GEN_Image) was produced, according to Eq. 8:

$$\mathrm{GEN\_Image(i,j)}=\sqrt{\left(\mathrm{GFIM}_{f,\theta,0^\circ}(i,j)\right)^2+\left(\mathrm{GFIM}_{f,\theta,-90^\circ}(i,j)\right)^2}. \tag{8}$$

Each point of the GEN_Image represents a measurement that is characterized as Gabor Energy [16].

Four (4) GEN_Images, corresponding to the filter orientations of θ = 0°, 45°, 90°, and 135°, were produced. From each GEN_Image, the following statistics were calculated as textural features and were used by the classification algorithms: mean value, variance, skewness, kurtosis, range and standard deviation.

Following multiple trials regarding the filter specifications, the best classification scores were achieved for textural features that were extracted from images that had been convolved with z × z Gabor filters, z = 5.

6.2 Generation of Gabor energy measure run length textural features

In the present study, new features are proposed based on the combination of GLRLM features and Gabor textural features.

These new features, labelled as Gabor Energy Measure Run Length (GEMRL) features, were extracted from each one of four Gabor Energy Images according to the following approach.

The Gabor Energy values of a Gabor Energy Image were transformed into the region 0–15 by means of a linear transformation providing a grey-level image of 16 discrete grey tones. Denoting this image as GEN_Image_θ_16 (where θ: 0, 45, 90 and 135° represents the orientation of the Gabor filter applied on the image), the new features were generated employing the Eqs. 9–13:

$$\mathrm{GEMRL1}=\frac{1}{P}\sum_{j=1}^{R}\frac{r_d(j)}{j^2}. \tag{9}$$

$$\mathrm{GEMRL2}=\frac{1}{P}\sum_{j=1}^{R}r_d(j)j^2. \tag{10}$$

$$\mathrm{GEMRL3}=\frac{1}{P}\sum_{i=0}^{G-1}[g_d(i)]^2. \tag{11}$$

$$\mathrm{GEMRL4}=\frac{1}{P}\sum_{j=1}^{R}[r_d(j)]^2. \tag{12}$$

$$\mathrm{GEMRL5}=\frac{1}{\mathrm{PN}}\sum_{\mathrm{j}=1}^{\mathrm{R}}\mathrm{r_d(j)}, \tag{13}$$

where, j represents the length of the run for the grey tone i, G and R are the numbers of grey tones and run-lengths in the GEN_Image_θ_16, respectively, PN is the number of pixels in the GEN_Image_θ_16, while the r_d, g_d, P are defined in the Eqs. 14–16:

$$r_d(j)=\sum_{i=0}^{G-1}q_d(i,j). \tag{14}$$

$$g_d(i)=\sum_{j=1}^{R}q_d(i,j). \tag{15}$$

$$P=\sum_{i=0}^{G-1}\sum_{j=1}^{R}q_d(i,j)=\sum_{i=0}^{G-1}g_d(i)=\sum_{j=1}^{R}r_d(j), \tag{16}$$

where, q_d (i,j) represents each element of the GLRLM computed along the angular direction d (d: 0, 45, 90 and 135°).

From each GEN_Image_θ_16, four GLRLM were calculated for the angular directions d of 0, 45, 90 and 135°. For each one of the GEMRL features, described by Eqs. 9–13, four values were extracted (one value from each GLRLM), as proposed by Galloway [14]. The mean of these four values was used as the final feature value [14].

References

1. Aigner T, McKenna L (2002) Molecular pathology and pathobiology of osteoarthritic cartilage. CMLS Cell Mol Life Sci 59:5–18
2. Altman R, Alarcón G, Appelrouth D, et al (1991) The American college of rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 34:505–514
3. Altman RD, Fries JF, Bloch DA, et al (1987) Radiographic assessment of progression in osteoarthritis. Arthritis Rheum 30:1214–1225
4. Atlamazoglou V, Yova D, Kavantzas N, Loukas S (2001) Texture analysis of fluorescence microscopic images of colonic tissue sections. Med Biol Eng Comput 39:145–151
5. Barandela R, Sánchez JS, Valdovinos RM (2003) New applications of ensembles of classifiers. Pattern Anal Appl 6:245–256
6. Bocchi L, Coppini G, De Dominicis R, Valli G (1997) Tissue characterization from X-ray images. Med Eng Phys 19:336–342
7. Boniatis I, Costaridou L, Cavouras D, Panagiotopoulos E, Panayiotakis G (2006) Quantitative assessment of hip osteoarthritis based on image texture analysis. Br J Radiol 79:232–238
8. Buckwalter A, Mankin HJ (1997) Instructional course lectures, the American Academy of Orthopaedic Surgeons — Articular Cartilage. Part II: degeneration and osteoarthrosis, repair, regeneration, and transplantation. J Bone Joint Surg Am 79:612–632
9. Campbell M, Machin D (1996) Medicals statistics, 2nd edn. Wiley Ltd., Chichester
10. Christodoulou CI, Pattichis CS, Kyriacou E, Pattichis MS, Pantziaris M, Nikolaides A (2005) Texture and morphological analysis of ultrasound images of the carotid plaque for the assessment of stroke. In: Costaridou L (eds) Medical image analysis methods. CRC Press, Taylor and Francis Group, Boca Raton, London, New York, Singapore, pp 87–135
11. Conrozier T, Tron AM, Balblanc JC, et al (1993) Measurement of the hip joint space using computerized image analysis. Rev Rhum Engl Ed 60:105–111
12. Daugman JG (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am 2:1160–1169
13. Efstathopoulos EP, Costaridou L, Kocsis O, Panayiotakis G (2001) A protocol-based evaluation of medical image digitizers. Br J Radiol 74:841–846
14. Galloway MM (1975) Texture analysis using gray level run lengths. Comput Graph Image Process 4:172–179
15. Gordon CL, Wu C, Peterfy CG, Duryea J, Klifa C, Genant HK (2001) Automated measurement of radiographic hip joint space width. Med Phys 28:267–277
16. Grigorescu SE, Petkov N, Kruizinga P (2002) Comparison of textural features based on Gabor filters. IEEE Trans Image Process 11:1160–1167
17. Ingvarsson T, Hägglund G, Lindberg H, Lohmander LS (2000) Assessment of primary hip osteoarthritis: comparison of radiographic methods using colon radiographs. Ann Rheum Dis 59:650–653
18. Jain AK, Duin RPW, Jianchang M (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal 22:4–37
19. Kellgren JH, Lawrence JS (1957) Radiological assessment of osteoarthrosis. Ann Rheum Dis 16:494–501
20. Lilliefors HW (1967) On the Kolmogorov–Smirnov-test for normality with mean and variance unknown. J Am Stat Assoc 62:399–402
21. Lumiscan 75, system specifications. Lumisys Inc. 1998; http://www.lumisys.com/support/manuals.html.
22. Martel–Pelletier J, Pelletier J-P (2003) Osteoarthritis: recent developments. Curr Opin Rheumatol 15:613–615
23. Ory PA (2003) Radiography in the assessment of musculoskeletal conditions. Best Pract Res Clin Rheumatol 17:495–512
24. Pizer SM, Amburn EOP, Austin JD, Cromartie R, Geselowitz A, Greer T (1987) Adaptive histogram equalization and its variations. CVGIP (Comput Vis Graph Image Process) 39:355–368
25. Radin EL, Rose RM (1986) Role of subchondral bone in the initiation and progression of cartilage damage. Clin Orthop 213:34–40
26. Recht MP, Goodwin DW, Winalski CS, White LM (2005) MRI of articular cartilage: revisiting current status and future directions. Am J Roentgenol 185:899–914
27. Sakellaropoulos P, Costaridou L, Panayiotakis G (1999) An image visualisation tool in mammography. Med Inform Internet Med 24:53–73
28. Sakellaropoulos P, Costaridou L, Panayiotakis G (2000) Using component technologies for web—based wavelet enhanced mammographic image visualization. Med Inform Internet Med 25:171–181
29. Specht DF (1990) Probabilistic neural networks. Neural Netw 3:109–118
30. Spector TD, Cooper C (1993) Radiographic assessment of osteoarthritis in population studies: whither Kellgren and Lawrence? Osteoarthr Cartil 1:203–206
31. Sun Y, Günther KP, Brenner H (1997) Reliability of radiographic grading of osteoarthritis of the hip and knee. Scand J Rheumatol 26:155–165
32. Theodoridis S, Koutroumbas K (2003) Pattern Recognition, 2nd edn. Elsevier Academic Press, Amsterdam, Boston, Heidelberg
33. Tourassi GD (1999) Journey toward computer-aided diagnosis: role of image texture analysis. Radiology 213:317–320
34. Tuceryan M, Jain AK (1998) Texture analysis. In: Chen CH, Pau LF, Wang PSP (eds) The handbook of pattern recognition and computer vision, 2nd edn. World Scientific Publishing Co., Singapore, pp 207–248
35. van Belle G, Fisher LD, Heagerty PJ, Lumley T (2004) Biostatistics. A methodology for the health sciences, 2nd edn. Wiley-Interscience, NJ