Combining GLCM with LBP features for knee osteoarthritis prediction: Data from the Osteoarthritis initiative

Khaled Harrar1,*, Khadidja Messaoudene2, Mohammed Ammar1

1LIST Laboratory, University M’Hamed Bougara of Boumerdes, Algeria
2LIMOSE Laboratory, University M’Hamed Bougara of Boumerdes, Algeria

Abstract

INTRODUCTION: Knee osteoarthritis is a chronic disease that can make a person more susceptible to develop health complications. It is a significant cause of disability among adults. In advanced stages, people can die from these complications.

OBJECTIVES: This paper introduces a quick and effective approach to classify knee X-ray images using LogitBoost and wavelet-based Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) to increase image classification accuracy and minimize training and testing time.

METHODS: The proposed technique involves image enhancement followed by Haar wavelet transformation. GLCM and LBP were extracted from the transformed image and these attributes were used to differentiate the radiographs into two groups of patients composed of 100 normal subjects (KL 0) and 100 pathological cases with osteoarthritis (KL 2). The validation of the classification was carried out using the K-fold cross-validation technique with k = 10.

RESULTS: The results revealed that the GLCM provided an accuracy of 77% and the LBP approach achieved an accuracy of 82.5%. Moreover, the combination of the two techniques LBP-GLCM improved the accuracy of the prediction with the LogitBoost model (91.16%). Compared to other classifiers (SVM, logistic regression, and decision tree), the LogitBoost provided a low root mean square error (RMSE) of 27.5%.

CONCLUSION: In addition, the proposed method was compared to the state-of-the-art and revealed the highest accuracy in the prediction of KOA, outperforming the methods existing in the literature.

Keywords: Knee osteoarthritis, X-ray images, DWT, GLCM, LBP, LogitBoost.

Received on 01 July 2021, accepted on 15 October 2021, published on 20 October 2021

Copyright © 2021 Khaled Harrar et al., licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license, which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.20-10-2021.171550

1. Introduction

Knee osteoarthritis (KOA) is one of the most dangerous arthritis in people over the age of 20. The global prevalence of KOA was 203 per 10000 person-years. In 2020, there are about 867 million people (20 years and older) with incident KOA worldwide [1]. Osteoarthritis (OA) is considered a disease of articular cartilage. It is usually due to several clinical and pathological disorders that result in structural and functional failure of synovial joints [2]. The risk factors for OA usually include age, gender, prior joint injury, obesity, genetic predisposition, or even mechanical factors such as misalignment and abnormal joint Shape [3]. KOA is graded into five stages according to the Kellgren and Lawrence system (KL) [4]. Table 1 shows the different grades of OA disease. The changes in the structure of the bone caused by OA are illustrated in Figure 1. The recognition of the precise stage of the disease is an essential part of avoiding complications of disability or death.
Multiple imaging modalities may be used to visualize OA. In daily clinical practice, the first most used imaging modality for initial diagnosis and routine follow-up is conventional radiography despite its known limitations. Hence, the necessity to develop automatic methods to increase the accuracy of the diagnosis.

KOA is one of the attractive topics for researchers to automatically detect and recognize this disease. Texture analysis is a suitable tool for characterizing images. It has been used in assessing bone fragility as osteoporosis [5,6]. KOA can also be detected using texture analysis from X-ray images. There are several methods in the literature for the detection of KOA. Janvier et al [7] combined the fractal analysis with a logistic regression model to examine whether trabecular bone texture (TBT) parameters assessed on computed radiographs could predict KOA progression. They achieved an AUC of 0.71 for Joint Space Narrowing (JSN) and 0.77 for JSN and TBT. In [8] the authors described a method that involved gathering additional texture information from the lateral and medial condyles of the distal femur. Their framework selected an optimal combination of different texture parameters from six different regions for evaluation with various classifiers. They achieved 72% of accuracy. Akter et al. [9] described a new approach to extract texture features in X-ray images for OA detection. The proposed method is based on Zernike orthogonal polynomials features and group method of data handling (GMDH) Neural Networks. The proposed method improved the detection accuracy by 72.5% for lateral images and by 68.3% for medical images.

Although the presented works provided promising results, they lack accuracy. This is the reason for developing new approaches to improve the accuracy of detection. The objective of our paper is to propose a new system for the detection and classification of KOA in knee radiographs. This work aims to apply the GLCM and LBP with Haar wavelet to discriminate two populations composed of 100 KOA (KL=2) patients and 100 healthy (KL=0) subjects. This is the first study combining GLCM and LBP for KOA detection. In our work, we have used the Hi parameter as the input feature of the LogitBoost classifier to discriminate the two populations. To avoid overfitting the k-fold cross-validation method was used to validate the results. Different classifiers including SVM, logistic regression, and decision tree are also tested and compared to our approach. The purpose of our work is to improve the performance of the screening of the disease. The main contributions of our work are:

1. Development of a system for KOA detection.
2. A combination of GLCM and LBP to improve the performance of the prediction.
3. Construction of different models to find the configuration providing the best performance.
4. Improve the performance of the classification of the patients, compared to the literature.

This paper is organized as follows: In Section 2, the dataset and the method are presented, including the DWT, features extraction, and classification. In section 3, the results obtained on the knee images are presented. Moreover, the analysis is discussed, and section 4 offers a conclusion.

![Figure 1. KL classification system for the evaluation of KOA severity [10]](image1)

<table>
<thead>
<tr>
<th>KL Grades</th>
<th>OA Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 0</td>
<td>No changes in the features of the X-Ray image (none)</td>
</tr>
<tr>
<td>Grade 1</td>
<td>Joint space narrowing (doubtful OA)</td>
</tr>
<tr>
<td>Grade 2</td>
<td>Definite joint space narrowing (mild OA)</td>
</tr>
<tr>
<td>Grade 3</td>
<td>Presence of multiple osteophytes, and some sclerosis (Moderate OA)</td>
</tr>
<tr>
<td>Grade 4</td>
<td>Marked joint space narrowing, large osteophyte (severe OA)</td>
</tr>
</tbody>
</table>

2. Materiel and methods

2.1. Dataset

The dataset used in our experiment was obtained from the publicly accessible osteoarthritis initiative database (OAI) [12]. The data contains 200 radiographs of the knee that have been categorized using the Kellgren and Lawrence rating system (KL0, KL2). We compared grade (no OA) to disease overall grade KL2 (mild OA).

![Figure 2. Knee radiographic image [13]](image2)
Combining GLCM with LBP features for knee osteoarthritis prediction: Data from the Osteoarthritis initiative

OAI data is a multicenter, longitudinal, and prospective observational study of KOA, composed of X-rays of the knee in fixed flexion [12]. In our study, we worked with the lateral region of the knee radiographic image. Figure 2 shows the region of interest (ROI) used in our work.

2.2. Methods

In this section, the techniques adopted in our approach for the classification of KOA are described. The proposed method consists of four steps which include a pre-processing step for artefact removing, a discrete wavelet transform step, features extraction using GLCM and LBP, and the classification into KL0 and KL2. Figure 3 shows the flow diagram of the suggested approach.

Preprocessing

Major sources of noise in an X-ray imaging system are inherent noise and X-ray quantum noise, which follow Gaussian distribution and the distribution of Poisson, respectively. The need for robust ways to eliminate noise and unwanted particles has therefore arisen [14]. In this work, the Gaussian filter is used for removing acquisition noise with kernels of 3×3 and 5×5. The choice of this filter is justified by the nature of the noise present in the images.

Figure 4 shows the results of filtering using the Gaussian filter.

Discrete Wavelet Transform (DWT)

Wavelets are widely used in image processing for denoising, edge detection, segmentation, compression, coding and decoding, texture analysis, etc. In this work, Haar wavelet transform was used for texture analysis. The Haar transformation can be defined as a sampling process in which an input data sequence with finer and finer resolution increase in powers of 2 is sampled by rows of the transform matrix. In image processing applications, the Haar transform provides a transform domain in which a type of differential energy is concentrated in localized regions [15].

Haar [16] defined a complete orthogonal function system in Lp ([0,1]), p ∈ [1, ∞] which take values from the set {0,2j: j ∈ N}. This function system has the property that any continuous function at intervals [0,1] can be presented by a uniform and convergent series in terms of this system’s elements. The Haar formulation is defined as follows [17,18]:

$$\text{haar}(0,t), \text{for } t \in [0,1], \text{haar}(1,t) = \begin{cases} 1, & t \in \left[0, \frac{1}{2}\right] \\ -1, & t \in \left[\frac{1}{2}, 1\right] \end{cases}$$  (1)

with 

$$\text{haar}(k,0) = \lim_{t \to 0} \text{haar}(k,t), \quad \text{haar}(k,1) = \lim_{t \to 1} \text{haar}(k,t)$$
and at the point of discontinuity within the interior (0,1)
\[ \text{haar}(k, t) = \frac{1}{2} (\text{haar}(k, t-0) + \text{haar}(k, t+0)) \].

### Feature extraction

#### A. Gray Level Co-occurrence Matrix (GLCM)

The GLCM approach is based on statistical studies of the pixel intensity distribution [19]. Haralick [20] proposed the gray levels matrices (GLCM) that have become one of the most commonly used and well-known texture characteristics. For a replacement \( P_d \) the matrix of the \( G^*G \) gray stage is defined as follows: the entry \((i, j)\) of \( P_d \) is the number of occurrences of the pair of gray levels \( i \) and \( j \) that are separated by a distance \( d \). It is formally described as [21]:

\[
P_d(i, j) = \sum \{ (r, s), (t, v): I(r, s) = i, I(t, v) = j \}
\]

Where \((t,v) = (r + dx, s + dy)\), \((r,s),(t,v) \in N \times N\) and \( \| \) is a set's cardinality.

Important features are taken out from the matrix as the texture representation. Five texture features are extracted: entropy, contrast, energy, correlation, and homogeneity derived from the co-occurrence matrix of the gray level (GLCM). The expressions of the texture features retained for this study are:

- **Energy**:
  \[
  \text{Energy} = \sum_{i,j} p(i, j)^2
  \]  

- **Contrast**:
  \[
  \text{Contrast} = \sum_{i,j} |i-j|^2 p(i, j)
  \]  

- **Correlation**:
  \[
  \text{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j}
  \]

- **Homogeneity**:
  \[
  \text{Homogeneity} = \sum_{i,j} \frac{p(i, j)}{1 + |i-j|}
  \]

- **Entropy**:
  \[
  \text{Entropy} = \sum_{i,j} p(i, j) \log p(i, j)
  \]

#### B. Local Binary Pattern (LBP)

Jala et al. [22] introduced local binary models as a way to summarize the local gray level structure. A valued local image pattern is created around every pixel. The initial LBP operator operates over an image in 3×3 windows. The pixels in these windows are thresholded by their center pixel value, multiplied by two forces in the direction of the clockwise or counter clockwise direction, and then summed up to achieve a pattern for center pixel The LBP operator formally takes the form:

\[
LBP(x_c, y_c) = \sum S(i_n - i_c)2^n
\]

Where \( i_n \) is the gray level of the pixel \((x_n, y_n)\), \( i_c \) is the gray level of the circular neighborhood of the pixel \((x_c, y_c)\), and \( S \) is the Heaviside function.

### Classification

In this study, the LogitBoost model was used to classify the input image into one of two grades of KOA: KL0 or KL2. LogitBoost is a new boosting technique that can handle multiclass problems by taking into account multiclass logistic loss [23]. This method has been used in several problems of classification and provided good results. We compared the classification performance of the proposed method to known classifiers (SVM, Logistic regression, and Decision Tree) based on AUC, accuracy, MCC, kappa coefficient, Precision, Recall, MCC, and RMSE.

### Statistical analysis

Several metrics were calculated in this study including:
- **TP**: the correctly defined number of true positive, pathological patients (KL0).
- **FP**: the number of false positives, pathological patients (KL2) incorrectly identified.
- **FN**: the number of false negatives, healthy patients (KL0) incorrectly identified.
- **TN**: the number of true negatives, healthy patients (KL0) correctly identified.

\[
\text{AUC} = \frac{TP + TN}{TP + FN + TN + FP}
\]

Kappa: The kappa statistic measures how precisely the instances identified by the machine learning classifier matched the data labeled as ground truth while adjusting for the performance of a random classifier as evaluated by anticipated accuracy. Kappa formula can be written as:

\[
\text{Kappa} = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FN)(FP + TN)(TP + FN)\cdot(FN + TN)}
\]

Precision: The precision assesses how accurate the model is in predicting positive labels. The formula for Precision is given by the expression (11):
Combining GLCM with LBP features for knee osteoarthritis prediction: Data from the Osteoarthritis initiative

\[ Precision = \frac{TP}{TP + FP} \]  

(11)

Recall: measurement of the percentage of positive actuals of a correctly defined model (True Positive). The formula for the recall is given by the expression (12):

\[ Recall = \frac{TP}{TP + FN} \]  

(12)

Matthews Correlation Coefficient (MCC): is used in computer vision to assess the quality of binary (two-class) classifications. It takes into consideration true and false positives and negatives and is typically regarded as a balanced metric that may be utilized even when the classes are greatly diverse in size.

\[ MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \]  

(13)

The Root Mean Square Error (RMSE) is a well-known method for calculating a model's error in estimating quantitative data. It is expressed mathematically as follows:

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \]  

(14)

3. Results and discussion

The proposed system is performed on various knee radiographs of different ages. The experiment is carried out on 200 Knee radiographs. Several attributes were calculated such as entropy, homogeneity, energy, contrast, and correlation. Texture parameters are derived from radiograph images and applied to several classifiers.

Figure 5 illustrates the performance of the prediction using various models and classifiers. In the first stage, we have tested the LBP features with DWT using four classifiers. In the second stage, we have tested GLCM features with DWT using four classifiers. In the last stages, we have combined LBP and GLCM using the four classifiers. The aim was to look for the configuration providing the best performance in terms of the prediction. From figure 5, the high rate of the classification was reached by LBP-GLCM and LogitBoost (ACC = 91.16%).

Based on the results obtained from the proposed methodology, the precision was better for the classification of LogitBoost. The data is trained and tested to obtain the classification rate for each classifier discussed above. The following Tables report the accuracy of the classification of the four classifiers and the other evaluation metrics were estimated. Experiments show that the combination of LBP and GLCM using LogitBoost classifier leads to a considerable improvement of the overall classification performance, with an AUC of 94.7% compared to the other classifiers (SVM, Logistic regression, and decision tree). The parameters of the classifiers are optimized to provide the best performance. Various parameters were included in this analysis to distinguish two groups composed of 100 subjects in KL0 and 100 patients in KL2.

Table 2 shows a comparison of the four algorithms using the LBP model. As can be seen, LogitBoost provided the best performance with an accuracy ACC = 82.5%, a precision of 83.1%, and an error of 37.88%. It outperformed the second-ranked method which is SVM with a remarkable margin. The worst performing approach was the decision tree technique with an ACC = 69.5%, and an error of 55.23%.

Table 3 shows the comparison between the classifiers using the GLCM method. As can be seen, the decision tree revealed the best performance in terms of the separation between the two populations based on all the metrics. This model achieved an accuracy of 77 %, followed by the logitBoost model with ACC = 74%. A precision of 83.5% and a Recall of 77%. The logistic regression provided less performance with 67% of accuracy, precision, and recall.

Table 4 illustrates the results of the combination of LBP and GLCM for KOA prediction. As can be seen, the performance of the classification was improved for all the classifiers, this is due to the fusion of advantages of the two techniques. Examining Table 4, the highest rate of the prediction was up to the logitBoost classifier with an accuracy of 91.16% and AUC of 94.7%. This model provided the lowest error (RMSE = 27.5%). This demonstrates the power of this classifier. The worst performing classifier was the logistic regression for this case with an accuracy of 65% and AUC of 68.2%, whereas the error was 57.75%.
### Table 2. Classification performance for LBP

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>ACC</th>
<th>Kappa</th>
<th>Precision</th>
<th>Recall</th>
<th>MCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogitBoost</td>
<td>0.856</td>
<td>0.825</td>
<td>0.6504</td>
<td>0.831</td>
<td>0.825</td>
<td>0.656</td>
<td>0.3788</td>
</tr>
<tr>
<td>SVM</td>
<td>0.792</td>
<td>0.79</td>
<td>0.5814</td>
<td>0.828</td>
<td>0.790</td>
<td>0.618</td>
<td>0.4583</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.726</td>
<td>0.675</td>
<td>0.3496</td>
<td>0.675</td>
<td>0.675</td>
<td>0.350</td>
<td>0.5381</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.694</td>
<td>0.695</td>
<td>0.3891</td>
<td>0.698</td>
<td>0.695</td>
<td>0.392</td>
<td>0.5523</td>
</tr>
</tbody>
</table>

### Table 3. Classification performance for GLCM

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>ACC</th>
<th>kappa</th>
<th>Precision</th>
<th>Recall</th>
<th>MCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogitBoost</td>
<td>0.789</td>
<td>0.74</td>
<td>0.48</td>
<td>0.757</td>
<td>0.740</td>
<td>0.497</td>
<td>0.413</td>
</tr>
<tr>
<td>SVM</td>
<td>0.725</td>
<td>0.725</td>
<td>0.45</td>
<td>0.738</td>
<td>0.725</td>
<td>0.462</td>
<td>0.5244</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.774</td>
<td>0.67</td>
<td>0.34</td>
<td>0.670</td>
<td>0.670</td>
<td>0.340</td>
<td>0.4479</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.757</td>
<td>0.77</td>
<td>0.54</td>
<td>0.835</td>
<td>0.770</td>
<td>0.601</td>
<td>0.3999</td>
</tr>
</tbody>
</table>

### Table 4. Classification performance for the combination of LBP-GLCM

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>ACC</th>
<th>kappa</th>
<th>Precision</th>
<th>Recall</th>
<th>MCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogitBoost</td>
<td>0.947</td>
<td>0.916</td>
<td>0.8057</td>
<td>0.911</td>
<td>0.912</td>
<td>0.806</td>
<td>0.275</td>
</tr>
<tr>
<td>SVM</td>
<td>0.795</td>
<td>0.795</td>
<td>0.59</td>
<td>0.836</td>
<td>0.795</td>
<td>0.630</td>
<td>0.4528</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.682</td>
<td>0.65</td>
<td>0.30</td>
<td>0.652</td>
<td>0.650</td>
<td>0.302</td>
<td>0.5775</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.778</td>
<td>0.805</td>
<td>0.61</td>
<td>0.829</td>
<td>0.805</td>
<td>0.634</td>
<td>0.4001</td>
</tr>
</tbody>
</table>

### Table 5: Comparison of the proposed method with the existing studies in the literature

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Method</th>
<th>Classifier</th>
<th>AUC (%)</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thomson et al [24]</td>
<td>2015</td>
<td>Fractal signature Shape information</td>
<td>Random forest</td>
<td>84.9</td>
<td>OAI</td>
</tr>
<tr>
<td>Kaggie et al [25]</td>
<td>2017</td>
<td>Texture parameters</td>
<td>Neural network</td>
<td>74</td>
<td>OAI</td>
</tr>
<tr>
<td>Brahim et al [26]</td>
<td>2019</td>
<td>Multivariate linear Regression (MLR)</td>
<td>Naive Bayes random forest</td>
<td>82.98</td>
<td>OAI</td>
</tr>
<tr>
<td>Proposed method</td>
<td>2021</td>
<td>DWT and LBP-GLCM</td>
<td>LogitBoost</td>
<td>94.7</td>
<td>OAI</td>
</tr>
</tbody>
</table>
Various studies have already shown the value of texture analysis in predicting the evolution of KOA. However, these works are difficult to compare since they use different techniques on small and non-free populations. Table 5 recapitulates the comparison of overall automated classification performances between the proposed method and other techniques for predicting KOA. The metric for the performance evaluation is AUC. Thomson et al. [25] demonstrated that combining the results of the shape and texture-based classifiers leads to considerable improvement in overall classification performance, with an AUC of 84.9%. Janvier et al. [7] estimated the AUC of the discrimination between grade KL0 (non-OA) and grade KL2 (minimum OA) and found 71%. Kaggies et al. [26] combined texture analysis with a neural network classifier to predict radiographic disease progression over 3 years, and achieved high sensitivity (86%), a specificity of 64%, and an AUC of 74% for the prediction of OA progression. Finally, Brahim et al. [27], proposed a computer-assisted method using Multivariate Linear Regression (MLR) and achieved an overall AUC of 82.98%. Table 5 revealed that our proposed method achieved the highest rate of AUC comparing to the other methods existing in the literature, with a rate of 94.7%. This high rate of classification is reached with the combination of the discrete wavelet decomposition GLCM, LBP, and logitBoost classifier which is suitable for medical applications. This work could add value to the screening of KOA in clinical routine and can help doctors in the decision of the prediction of the disease.

4. Conclusion

In this paper, an application of the Local Binary Pattern combined with Gray Level Co-occurrence Matrix using LogitBoost classifier was presented to discriminate the knee osteoarthritis and non-knee osteoarthritis X-ray images. The texture parameters are computed from the extracted ROIs in the osteoarthritis initiative dataset. The results revealed good performance for the combination of the LBP-GLCM operators using the k-fold Cross-validation (k-fold =10) method with an accuracy of 91.16%. Experimental results suggest that more than 91% of KOA was diagnosed. The results achieved are promising when compared to methods existing in the literature. In the future, we plan to improve the classification of knee OA by using different statistical features and augmenting the dataset. It is interesting to extend the classification to differentiate between the other KL classes, to assess the consistency of each grade of osteoarthritis.

References


