A Comprehensive Digital Knee X-ray Image Dataset for the Assessment of Osteoarthritis

Shivanand S. Gornale*, Pooja U. Patravali and Prakash S. Hiremath
Department of Computer Science, Rani Chanamma University, India

Abstract

Osteoarthritis is a chronic degenerative joint disease called as “Wear and Tear” arthritis that results in progressive erosion of articular cartilage and results in disability. Earlier work reveals that various segmentation and classification methods have been implemented for the assessment of the disease but due to unavailability of standard benchmark database the comparative analysis of individual methods is still missing. In this paper, we present a comprehensive dataset of digital Knee X-ray images complaint with DICOM (Digital Imaging and Communications in Medicine) standards with manual annotations made by two different medical experts as per Kellgren and Lawrence grading system. The paper divulge the study of various state of art methods that includes region based methods, density based methods and statistical evaluation measures for early assessment of ailment and severity level as per KL grading system against this dataset.

INTRODUCTION

Osteoarthritis is one of the most common forms of arthritis disease that is seen mostly in females, overweight and elderly people. Osteoarthritis (OA), is a joint disease that mainly affects cartilage. Osteoarthritis is a chronic degenerative joint disease called as “Wear and Tear” arthritis that results in progressive erosion of articular cartilage [1]. Here joint lining becomes pitted, eroded; uneven and painful. The day to day activities get impeded due to severe pain and restricted movement. Cartilage is the protective connective tissue that covers the end of bones in a joint. Healthy cartilage allows easy glide of bone in the joint and prevents them from rubbing each other [2]. In Osteoarthritis the top layer of cartilage breaks down and wears away. This allows the bones to rub each other causing pain. It commonly affects the joint in the knee, hip, spine and feet [3]. It is observed that most of the people who are affected by Osteoarthritis are older than age 45 and women are more susceptible to the disease than men [4]. The normal knee and Osteoarthritic knee with respect to X-ray imaging is depicted in Figure 1(a) and 1(b).

Based on the radiological parameter of knee x-ray the extremity of the disease can be defined using KL grading system. Kellgren and Lawrence (KL) have developed a standard grading system to understand the severity of Knee Osteoarthritis [5,3]. The clinical reviews have insisted that there is strong correlation between Knee injury and Osteoarthritis [6]. The Kellgren and Lawrence system is most common method used to classify the individual joints into 5 different grades essentially to discern the severity of the disease. The severity level as per KL grading system, with description is predicted in Table 1.

In this paper, we present a self created 1650 digital X-ray images of knee joint with DICOM standards which are collected from well reputed hospitals and diagnostic centres. Each radiographic knee image is manually annotated /labelled as per Kellgren and Lawrence grades by 2 medical experts. The two experts are well experienced orthopaedic surgeons who examine 70 to 100 radiographic images every day. The various segmentation and classification methods have been evaluated on the dataset and the results are also presented.

MATERIALS AND METHODS

In this section, we summarize the existing datasets for digital Knee x-ray for evaluation of OA and present a detailed description of the presented dataset.

Existing Database

The earlier work revealed that most of the researchers have created their own private dataset for the disease assessment. It is observed that there are relatively few digital knee x-ray image dataset publically available for Osteoarthritis assessment. The two databases that are available are Osteoarthritis Initiative (OAI), and Multicenter Osteoarthritis Study (MOST).

The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01- AR-2-2259; N01 AR-2- 2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National...
Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. The baseline cohort of the OAI dataset contains MRI and X-ray images of 4,476 participants. From this entire cohort, 4,446 X-ray images are labelled according to KL grades for both knees as per the assessments by Boston University X-ray reading centre (BU), and the remaining x-ray images stay unlabelled [7].

MOST is comprised of four cooperative grants (Felson – AG18820; Torner– AG18832; Lewis – AG18947; and Nevitt – AG19069), funded by the National Institutes of Health, a branch of the Department of Health and Human Services and conducted by MOST study investigators. This manuscript was prepared using MOST data and does not necessarily reflect the opinions or views of MOST investigators. The MOST dataset includes lateral knee radiograph assessments of 3,026 participants. From this, 2,920 radiographs have been assigned KL grades for both knees as per the assessments by Boston University X-ray reading centre (BU), and the remaining x-ray images stay unlabelled [7].

To request access to these databases, we should be sponsored by an NIH (National Institutes of Health), recognized institution with a Federalwide Assurance and have a research related need to access NDA (NIMH (National Institute of Mental Health) Data Archive), data. NDA federated repositories have their own access requirements. Therefore, it becomes difficult for accessing such standard databases by unauthorised users or common researcher. Apart from this, these databases comprise of X-ray as well as MRI images in different views (Anterior post view, lateral view etc), where in some images remain unlabelled. In addition to this, how many medical experts have participated in annotating these X-ray images is also lacking in the existing public-privately available databases.

Subsequently, this stimulated us to create our own private database by visiting well reputed hospitals and diagnostic centres exclusively containing anterior posterior view of knee joint which are annotated by two distinctive well experienced medical experts.

**Image Acquisition**

Image acquisition can be designated as the action of fetching an image from some source which is further processed to get new and better image. The images are generally created by combining the illuminations of sources that are instigated by some X-ray producing sources, radar, ultrasound energy sources etc. It can also be done by absorbing the energy generated by the elements of a view or picture [9]. The process of Image acquisition is completely reliant on hardware system that consists of sensors which in turn is a hardware device. The sensors are meant to measure the reflected energy generated by the elements of an image and convert them into electrical charges. Some of the common devices used for retrieval of images are high resolution camera, x-ray machines, 2-D charged coupled device (CCD) camera etc. For the experimentations the fixed-flexion digital knee X-rays were acquired using a PROTEC PRS 500E X-ray machine.

**Ground Truth Collection**

From the literature survey, it is found that some standard
databases which are available consists of Radiographic as well as Magnetic Resonance Imaging (MRI) images of human joints like hip, hand, finger, spine etc. It is observed that very less numbers of digital knee X-ray images are available with no medical experts labelling or annotations according to KL grading framework. This motivated us to create our own dataset of digital Knee X-ray images, to carry out the proposed research work. The fixed-flexion digital knee X-rays were acquired using a PROTEC PRS 500E X-ray machine. Original images were 8-bit 1350x2455 grayscale Digital Imaging and Communications in Medicine (DICOM) images. DICOM standard is international standard to transmit, store, retrieve, print, process and display medical imaging information. It makes information interoperable and also meets the evolving technologies and requirements of medical imaging.

To initiate the research, 1650 digital X-ray images of knee joint with DICOM standards are considered which are collected from various hospitals and diagnostic centres. Each radiographic knee image is manually annotated /labelled as per Kellgren and Lawrence grades by 2 medical experts. The two experts are well experienced orthopaedic surgeons who examine 70 to 100 radiographic images every day. For all 1650 images, the grade-wise annotations made by two ortho-surgeons as per KL grading framework are summarized in Table 2.

Table 2: Grade-wise annotations by two different medical experts.

<table>
<thead>
<tr>
<th>KL Grade</th>
<th>Medical expert-I</th>
<th>Medical expert-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal(G-0)</td>
<td>514</td>
<td>503</td>
</tr>
<tr>
<td>Doubtful(G-1)</td>
<td>477</td>
<td>488</td>
</tr>
<tr>
<td>Mild(G-2)</td>
<td>232</td>
<td>232</td>
</tr>
<tr>
<td>Moderate(G-3)</td>
<td>221</td>
<td>221</td>
</tr>
<tr>
<td>Severe(G-4)</td>
<td>206</td>
<td>206</td>
</tr>
<tr>
<td>Total</td>
<td>1650</td>
<td>1650</td>
</tr>
</tbody>
</table>

Segmentation methods

Segmentation is one of the crucial steps in image processing; the entire analysis becomes uncomplicated if this step is done appropriately. From the literature we came across that most of the automatic segmentation algorithms are partially successful in terms of extracting region of interest. The perceptible data can be further mined from the segmented image for disease analysis [10]. The exploration of Osteoarthritis is implemented through distinct segmentation techniques which may further help for the appropriate and clear diagnose of the disease grade wise (KL grading system). In our research work we have implemented appropriate segmentation methods/techniques to extract the cartilage region for the ease analysis of ailment. The self created dataset is assessed using different image segmentation techniques that includes Edge based, Texture based, Otsu’s based, Active contour based and density based method.

Initially the dataset was utilised to extract the cartilage region using active contour method [11,12], and later to extract the object boundaries of cartilage, different edge operators, texture based method and Otsu’s method were implemented [13]. Primarily all the edge operators were utilised but among them Sobel operator and Prewitt operators outcomes were prevailing compared to other operators [11]. Texture based segmentation is generally used to examine the gray level transitions in foreground and background of medical images. In the experimentation, Texture method discriminated the texture of cartilage region by employing statistical estimations [12]. However Otsu method basically works on of global thresholding which particularly depends on the gray values of images. The optimal gray level threshold value is chosen to separate an object of interest in an image based on the threshold value [14,15].

Further as a novel approach we developed an algorithm that automatically identify and extract the cartilage region based on the density of pixels. Firstly to identify the cartilage region the bone edges are detected and multiple segments are formed based on the density of pixels. Subsequently the area of each segment is computed and the segment with the highest area is extracted [16].

Computation

In feature extraction the most significant features have to be figured out to understand the patterns of segmented region so that the task of classifying the pattern is made easy by a formal procedure [9,17]. The cartilage region which was segmented using different segmentation methods was further assessed using multiple feature extraction methods. The diverse global and local features were computed for appropriate classification. The computation includes assembling, formulating, determination & clarification of data by measuring the similarities between shapes represented by their features. It also contained data about how image pixels intensities with a specific position in association to one another occurring together. The objective of feature computation is to identify the texture by figuring it locally at each pixel area and exhibiting the subsequent codes for appropriate classification [18,19].

Classification

The classification is the process of automatically categorizing the computed features of the images into groups or classes accordingly. For a specific application image classification is implemented using a computer program which is referred as classifiers [20]. Classification process is generally carried out in two different methods; supervised classification and unsupervised classification. In supervised methods we train the machine using data which are well labelled or annotated to make predictions. Where as in unsupervised method no any labelling are used and we make the machine to learn on its own to discover information. In our work we are concentrating on supervised classification algorithms as we already have a dataset which is been manually labelled by two different medical experts [21,22].

EVALUATION MEASURES

Three different measures are proposed for the quantitative evaluation of the ailment. The measures includes; 3.1) identification and extraction of cartilage region 3.2) segmentation of cartilage region 3.3) estimation of cartilage thickness.

Identification and extraction of cartilage Region

Identification of region of interest of an image can be implemented by discovering and partitioning the entire
image into significant structures. Further it is segregated from background or foreground which can be scaled, estimated or evaluated for processing\[9,23\]. The entire identification process is divided into four main steps. Firstly detection of bone edges is implemented by finding the difference with the neighboring pixels. The bone edge detection at pixel (x, y) for input image \( I \) with respect to row and column is given in (1) and (2).

\[
I(x, y) = I(x, y) - I(x-1, y) \quad (1)
\]

\[
I(x, y) = I(x, y) - I(x, y-1) \quad (2)
\]

After detecting the bone edges’ pre-processing of the image is conducted that includes noise removal using adaptive median filter. Mainly the images acquired from x-ray detectors are eminent in salt and pepper noise. Adaptive median filtering is helpful in lessening salt and pepper noise or impulsive noise and beneficial in preserving edges in an image that result in reduction of irregular noise \[24\]. After identification we go for extracting the region of interest (ROI) by segmenting the image into multiple parts along the axis. The area of each segmented part is estimated to detect the bone density. The region in the image with high pixel value is always more dense or thicker, which results in high density value. Based on this density value, identifying the region of interest becomes easier \[25\].

The final step incorporates enhancing the region with high density value using sine adaptive filter. The steps used by sine adaptive filter are given in equations (3-6).

\[
C_f = 0 : \frac{\Pi}{2} = 0 \quad (3)
\]

\[
A_f = \frac{\Pi}{2} = F_w \frac{\Pi}{2} \quad (4)
\]

\[
C_f (C_f > A_f) = \frac{\Pi}{2} \quad (5)
\]

\[
I_{\text{area}} = I_{\text{area}} \sin(C_f) \quad (6)
\]

where, \( C_f \) is a filter co-efficient and \( F_w \) is a Filter Width calculated based on the X-ray reconstructed image characteristics. Images recorded from x-ray detectors always have the data in the centre portion of the image \[26\]. Therefore by defining the filter co-efficient as a sine wave, we are adding more weight to the data in the middle of the image. Adaptive sine filter is used for allocating weights based on the geometrical axis of the reconstructed image. Thus the region of interest is accurately extracted that can be further used by active contour algorithm for segmentation \[16,27\].

**Segmentation of Cartilage region**

The cartilage region is further segmented using active contour algorithm. The extracted ROI is again used as one of the input parameter to active contour algorithm along with 3x3 masks. The algorithm is utilized to fragment an item by giving it a chance to settle much like a contracting snake around the cartilage region \[10\][28]. We specify the curves on the image that move to find object boundaries in the form of mask. Mask is a binary image that specifies the initial state of active contour algorithm and evolves iteratively to find the object boundaries \[29\].

**Estimation of cartilage thickness**

The cartilage thickness is computed using basic shape features which are calculated based on connected components stored in contiguous and dis-contiguous regions. The four main shape features used for computation of cartilage thickness are \[30\]:

- **Area:** It is the total number of on pixels in the image region.
- **Major axis length:** It indicates the length of major axis of an object in pixels. Major axis length is given in equation 7, Where \( a \), \( b \) are distance from each focus point.
- **Minor axis length:** It signifies the length of minor axis of an object in pixels. Minor axis length is given in equation 8, Where, \( m \) is distance between focal point, \( x \) and \( y \) are distance from each focal point.
- **Perimeter:** It is number of boundary pixels i.e. it calculates the distance between each adjoining pair of pixels around the border of the region.

**BENCHMARK RESULTS**

The self-created dataset has been used to assess some of the algorithms we have proposed earlier. These methods are summarized and the results are provided for benchmarking purposes.

For benchmarking, image dataset of 1650 images is assessed to carry out the experimentation. Here 50% of images are used for training and 50% of images are used for testing. This process is repeated 5 different times and the average accuracies are considered.

The implementation was performed using decision tree and K-NN classifiers. The aggregate accuracies obtained utilizing two classifiers are given in Table 3 below:

From Table 3 it is observed that K-Nearest Neighbour has outperformed compared to decision tree. In order to interpret the details of classification accuracy, confusion matrix of K-NN and decision tree classifier are given in Table 4 & Table 5.

In this work, along with the classification accuracy, the performance of the proposed method is evaluated with the following matrices like Precision, Recall and F-measure. The precision recall and F-measure are given in equation (9-11).

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (9)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (10)
\]

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Aggregate Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>99.81%</td>
</tr>
<tr>
<td>Decision tree</td>
<td>95.0%</td>
</tr>
</tbody>
</table>

\[
\text{Table 3: Aggregate Accuracies using two classifiers.}
\]
Table 4: Confusion Matrix of classification by K-NN classifier.

<table>
<thead>
<tr>
<th>Class</th>
<th>Normal</th>
<th>Doubtful</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (G0)</td>
<td>512</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doubtful (G1)</td>
<td>2</td>
<td>477</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mild (G2)</td>
<td>0</td>
<td>0</td>
<td>232</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Moderate (G3)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td>Severe (G4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>206</td>
</tr>
</tbody>
</table>

Table 5: Confusion Matrix of classification by Decision tree classifier.

<table>
<thead>
<tr>
<th>Class</th>
<th>Normal</th>
<th>Doubtful</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (G0)</td>
<td>468</td>
<td>18</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doubtful (G1)</td>
<td>36</td>
<td>447</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mild (G2)</td>
<td>10</td>
<td>10</td>
<td>228</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Moderate (G3)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td>Severe (G4)</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>206</td>
</tr>
</tbody>
</table>

\[
F_{\text{Measure}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)
\]

here, TP stands for True Positive, FP stands for False Positive and FN stands for False Negative. The precision, recall and F-measure for both the classifier are depicted in table 6.

The thickness of cartilage computed using shape descriptor according to KL grading system is given in table 7 and graphical representation is depicted in Figure 3.

STATISTICAL TEST OF SIGNIFICANCE

In research, more often subjective interpretations are not enough in declaring the results, for that statistical evidence is needed. To exhibit the statistical verification, significance test is fundamentally performed to regulate whether the outcomes of dataset are statistically significant or not. In our research work two different tests namely Chi-Square test and T-test are performed on all the experiments mainly to check the statistical test of significance [15,16,19,21,29,30]. Chi-Square test basically allows to test whether there is relationship between two variables or no. In this test null hypothesis is about relationship existing between two variables. The test is performed between the results of all proposed algorithms and the manual annotation made by medical expert-I and expert-II. Whereas T- test is meant to answer the question whether two groups are statistically different from each other. In this test, a null hypothesis is about mean difference between two groups is zero or some value. This test is basically done to check difference in the opinions of two experts. When these tests were performed on all the experiments it demonstrated that there exists a strong correlation between results of proposed algorithm and the annotations made by both medical experts and also represented that difference in opinions

Table 6: Performance measures of two classifiers.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>0.9983</td>
<td>0.9983</td>
<td>0.9983</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.9652</td>
<td>0.9577</td>
<td>0.9614</td>
</tr>
</tbody>
</table>

Table 7: Mean and Standard deviation of Cartilage thickness grade wise.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Computed Cartilage thickness (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal-Grade0 (514)</td>
<td>6.332±1.309</td>
</tr>
<tr>
<td>Doubtful-Grade1 (477)</td>
<td>5.274±1.002</td>
</tr>
<tr>
<td>Mild-Grade2 (232)</td>
<td>4.469±0.889</td>
</tr>
<tr>
<td>Moderate-Grade3 (221)</td>
<td>3.289±0.579</td>
</tr>
<tr>
<td>Severe-Grade4 (206)</td>
<td>2.407±0.305</td>
</tr>
</tbody>
</table>

Figure 2 Weight bearing Anterior Posterior View of Knee Joint.

Figure 3 Graphical representation of cartilage thickness.
of two experts is due to random variations in samples and not due to experts. From both the statistical tests, the proposed algorithms are found to be significantly effective and can provide computer-aided assistance to doctors for rapid assessment of the knee OA severity level from knee X-ray images so that the patients get better and accurate timely treatment.

CONCLUSION

This paper presents a dataset of digital Knee X-ray images for the evaluation of Knee Osteoarthritis. The ground truth derived from experts is comprehensive, providing manual annotations according to KL grading system. The dataset is intended to foster active research in image analysis and contribute towards developing robust solutions for automated Knee Osteoarthritis assessment.

ACKNOWLEDGEMENT

Authors would like to thank Department of Science and Technology (DST) for financial assistance under Women Scientist-B Scheme. (Ref No: SR/WOSB/65/2016(G)). Authors are also thankful to Dr. Chetan M. Umarani, Orthopaedic Surgeon, Gokak-Karnataka for providing knee X-ray images and Dr. Kiran S.Marathe, Orthopaedic Surgeon, JSS Hospital, Mysore, Karnataka-India for validation of the computed results by visual inspection.

REFERENCES


8. http://most.ucsf.edu/


