A Comparative Study of Texture Analysis Techniques for Osteoarthritis Classification Using Knee X-ray Imagery.

Sophal Chan[a], Kwankamon Dittakan[a]

[a] *College of Computing, Prince of Songkla University, 80, Moo 1, Vichitsongkram Rd, Kathu, Kathu, Phuket, Thailand 83120*

Abstract

Knee Osteoarthritis (OA) is one of the most famous diseases in aging society, it is affected million people around the world and affected over 10 million people in Thailand. Early detection is the most important way to prevent knee OA get into the serious stage of OA. The early detection of OA can be applied to medical imaging include x-ray, MRI, and CT. In the research study is focused on early detection of OA by applying image processing and classification technique to knee x-ray imagery. Texture analysis is presented in the research for analyzing the properties of the bone surface into four regions of interests (ROI). Texture feature can be analysis by texture descriptor for instance: : i) Histogram feature, ii) Local Binary Pattern (LBP), iii) Completed LBP (CLBP), iv) Rotated LBP (RLBP), v) LBP Histogram Fourier (LBP-HF), vi) LBP Rotation Invariant (LBP_ri), vii) Gabor, viii) Haralick, ix) Local Configuration Pattern (LCP) and x) Local Ternary Pattern (LTP). The next process after analyze texture is feature selection, feature selection is used to reduce the dimensionality of feature include: i) Correlation-based Feature Selection (CFS), ii) Chi-square, iii) Gain Ration, iv) Information Gain and v) Relief. Lastly, the classification process have applied nine different machine learning algorithms as follows: i) Decision Tree(C 4.5), ii) Decision with binary true, iii) Average One-Dependence Esti-

*Corresponding author

Email addresses: `sophal.c@phuket.psu.ac.th` (Sophal Chan), `kwankamon.d@phuket.psu.ac.th` (Kwankamon Dittakan)

mators (AODE), iv) Bayesian Network, v) Nave Bayesian Classier, vi) Support Vector Machine (SVM), vii) Logistic, viii) Sequential Minimal Optimization, and ix) The backpropagation algorithm. As a result, the four major propose are shown: i) the best result of ROI went to Literal Femur(LF), ii) LBP become the best descriptor amount of nine in the research study, iii) The Best feature selection in the research is CFS, and iv) The Best learning algorithm for the research study is represented by Bayes network. The measurable parameter of each result was measured by Area Under Curve (AUC), Accuracy (AC), Sensitivity (SN), Specicity (SP), Precision (PR), and F-Measure.

Keywords: Texture Analysis, Osteoarthritis, Knee OA, Image Classification.

---

1. Introduction

Osteoarthritis called in short as OA is the degenerative joint disease which happens in human joints. OA is the most prevalent disease in the aging society of the joints and the most common disease of arthritis which happening to millions of people in the United States [1], while in Thailand has almost reach10 million people are affected by OA or consider as 13% of all population in Thailand, based on the data of National Statistical Office (NSO) in 2014. With the respect NSO, Thai is of the country which is going to be the aging society country. In the aging society, knee OA has found out happen on men is approximately 10% and 13% in women [2]. Symptomatic of knee OA can be detected by the presence of pain, swelling, stiffness in the knee, reduce the ability of movement and head the cracking sound when the knee getting a move. Furthermore, the OA can get the early detection by the medical image to prevent the OA get into the serious condition. Medical imaging is widely used for OA early detection include: X-ray image, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI).

In the research study aims to apply image processing on medical x-ray image to detect OA. Image processing in the scientific field which is studied and analyze the digital images, for example, Medical image, satellite image, etc. to produce a better image. Image processing is a technology to study any algorithm to enhance the image or extract some useful features from the image to study for any specific purpose include: biology, medicine, astronomy, biometric and so on. In other words, image processing is specific technology use for i) classification, ii) feature extraction, iii) multi-scale signal analysis, iv) pattern recognition and v) projection. On the other hand, the implementation of image processing for classification in medical x-ray images is proposed in this work.

With the research motivation is focused on classification of the OA and non-OA x-ray image from analyzing the specific region of knee (Region Of Interest: ROI) which divided into four ROI: i) Literal Femur(FM), ii) Literal Tibia(LT), iii) Medial Femur(MF), and iv) Medial Tibia(MT). The research is aimed to classify images by applying image processing classification techniques. In medical image classification techniques are divided into two feature analyzing category: i) image classification by texture feature analysis and ii) image classification by shape feature analysis. In addition, the texture feature analysis is used to classify the OA or non-OA in the research while there are 10 techniques of texture analysis are used include: i) The first level of Gray-Level Co-Occurrence Matrix (GLCM), ii) Local Binary Pattern (LBP), iii) Completed LBP (CLBP), iv) Rotated LBP (RLBP), v) LBP Histogram Fourier (LBP_HF), vi) LBP Rotation Invariant (LBP_ri), vii) Gabor, viii) Haralick, ix) Local Configuration Pattern (LCP) and x) Local Ternary Pattern (LTP). Next, the ten texture feature descriptor is applied with five different feature selection include: i) Correlation-based Feature Selection (CFS), ii) Chi-square, iii) Gain Ration, iv) Information Gain and v) Relief. Lastly, the learning algorithms are applied to get the final result of the work, learning algorithm is presented in the work include: i) Decision Tree(C 4.5), ii) Decision with binary true, iii) Average One-Dependence Estimators (AODE), iv) Bayesian Network, v) Nave Bayesian Classifier, vi) Support Vector Machine (SVM), vii) Logistic, viii) Sequential Minimal Optimization, and ix) The backpropagation algorithm.

The rest of the article is presented related work is section II, Proposed of the research framework is discussed in section III. The brief information of each texture feature descriptors is presented in section iv. The discussing of feature selection and classification by learning algorithms are described in section v. Section vi is presented data collection and evaluation is discussed in section vii. Finally, the conclusion, acknowledgment, and references are presented in section viii, ix, and x particularly.

2. Related work

In recent, the OA detection and classification have been focused widely [3, 4, 5, 6, 7]. For OA early detection can apply with medical imaging and professional clinician to classify the OA. On the other hand, the implementation of classification techniques to medical imaging for OA detection has been spreading as an interesting topic in image processing research fields. The texture is one of the useful solutions in medical image processing for diagnosis and detection OA in clinical [8, 9] . The texture is one of the most import properties in images and shows the arrangement of pixels in objects to analyze. In work [3], the research focused on ROI of Tibia texture to analysis the OA. Texture analysis can apply various of medical images including: x-ray [3], MRI [10], CT, and Infrared [4]. In addition, texture in an object can be analysis by texture descriptor which is a technique to descript the texture into numeric which include: LBP [8, 11, 12], CLBP [13],LBP_ri [14],LBP-hf[15], LTP [16, 17], and etc. For an instant, a research study [6] introduced Haralick feature amount 4 features to analyze Knee OA x-ray image.

The research study [6] have detection knee OA by divided knee x-ray into nine blocks then the block 4,5 and 6 consider as the ROI of analysis texture as GLCM. In the work has focused on features of GLCM include: i) skewness, ii) kurtosis, iii) standard deviation, and iv) energy. The result of the work is received from SVM classifier.

With the present of statistic feature include: i) entropy, ii) mean, iii) median, iv) standard deviation and Tamura texture feature are implemented to analyze texture for OA radiographic in [18].

3. Proposed Framework

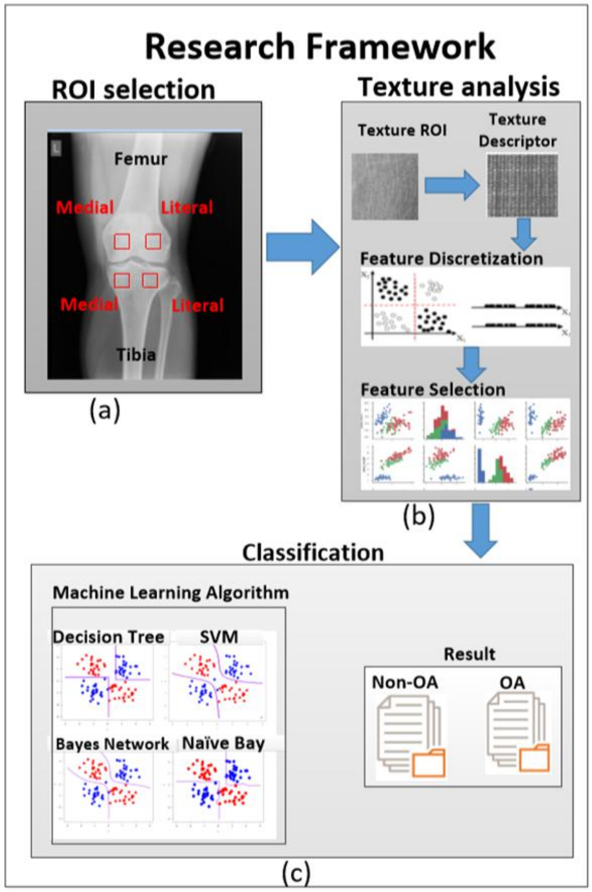

Figure 1: Texture Analysis on Knee OA Classification Techniques

With the respect to the framework in Fig.1, the framework has divided into three main processes, include: (a) Region of interest (ROI) selection, b) Texture Analysis, and iii) Classification process. First and foremost, the ROI selection process is the process to select the specific area which is called the region of interest, the purpose for this process in to select only the area which is considered to have the unique identity to detect the texture. In addition, the ROIs in this process are selected from 4 different area as shown in Fig.1 (a), the four ROI include two ROIs in the femur bone on the lateral and medial side, while the other two in the tibia on lateral and medial side. The final output of the ROI selection process is the four ROI, include: i) Femur Medial ROI, ii) Femur Lateral ROI, iii) Tibia Medial ROI, and iv) Tibia Lateral ROI, the final output is used as the input of the (b) Texture Analysis process. In the texture process has separate into four sub-processes: i) Choose ROI, ii) Apply the texture descriptor on ROI, iii) Feature discretization, and iv) Feature selection.

For the first sub-process, the ROI can be selected one among four, then the selected ROI is used for extraction the feature which is called feature descriptor. The feature descriptor is mention more in section IV. The feature discretization process is used to group the feature which is shared the same identity. In the research, has grouped the feature in 10 bins. The Feature is next process after finishing the discretization process. The purpose of feature selection is to select the useful feature for the classification process. In the feature selection process, the 5 feature selection techniques are applied in the research, for instance: i) Correlation-based Feature Selection (CFS), ii) Chi-square, iii) Gain Ration, iv) Information Gain and v) Relief. Finally, the classification process is used to classify the texture which has OA or non-OA. In the classification process, there are 10 machine learning algorithms are used to classify for OA and non-OA case, while the evaluation is measured by Area Under Curve (AUC), Accuracy (AC), Sensitivity (SN), Specificity (SP), Precision (PR), and F-Measure. The Machine learning algorithm is discussed more in section V. The texture descriptor is discussed in section IV.

4. Texture Descriptor

Texture descriptor is one of the most important techniques to classify the similarity image. There are ten feature descriptors are used in the research work include:

(1) The histogram features

The histogram feature of the grey level image is received by state of the art histogram based feature, include:

- Mean

$$\mu = \sum_{i=1}^{N} (iP(i)) \tag{1}$$

Where P(i) is the probability distribution of bin i, which P(i) can be written as:

$$P(i) = \frac{H(i)}{M} \tag{2}$$

H(i) is the histogram function and M is the number of blocks.

- Variance

$$\sigma^2 = \sum_{i=1}^{N} (i-\mu)^2 P(i) \tag{3}$$

- Skewness

To define the Skewness equation, the standard deviation have to be found first:

$$\sigma = \sqrt{\sum_{i=1}^{N} (i-\mu)^2 P(i)} \tag{4}$$

With the respect to equation (4), the skewness can be defined as:

$$skew = \frac{1}{\sigma^3} \sum_{i=1}^{N} (i-\mu)^3 P(i) \tag{5}$$

- Kurtosis

$$Kurtosis = \frac{1}{\sigma^4} \sum_{i=1}^{N} (i-\mu)^4 P(i) \tag{6}$$

- Energy

$$Energy = \sum_{i=1}^{N} [P(i)]^2 \tag{7}$$

- Entropy

$$Entropy = -\sum_{i=1}^{N} P(i) log_2[P(i)] \tag{8}$$

(2) **Local Binary Pattern (LBP)**

Local Binary Pattern (LBP) [19] is used to label the pixel which applied thresholding the neighborhood of each pixel with the output as the binary

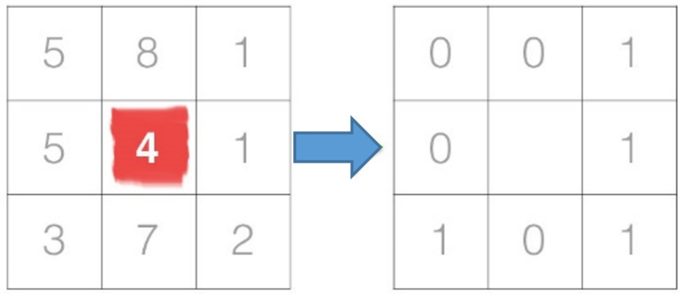

Figure 2: LBP Operator.

number. The basic of LBP operator is shown in Fig.2 below:

In addition, LBP at pixel (x_c, y_c) can be calculated by the equation below:

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} S(i_P - i_c)2^P \tag{9}$$

Where

P is the pixels surround in the circle neighborhood.

R is a radius of circle.

Ic and ip are the grave-level values of the center point.

s(x) is a function which is represented as:

$$s(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \tag{10}$$

Beside the LBP, there is another Completed LBP which is use the basic of LBP called CLBP.

(3) **Completed LBP (CLBP)**

Completed LBP or CLBP [13], a local region is defined by center pixel and a local difference sign-magnitude transform (LDSMT). In the research study is focused on LDSMT, LDSM T breaks down the image local structure into two component: the difference signs (CLBP_S) and the difference magnitudes (CLBP_M). The implementation of CLBP_S and CLBP_M are shown in the Fig.3 below:

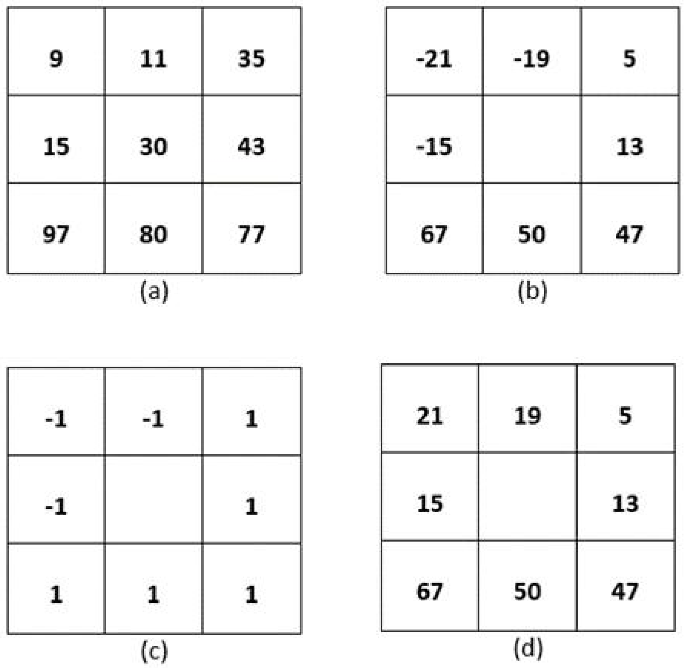

Figure 3: (a) 3x3 pixel, (b) the local different, (c) CLBP_S and (d) CLBP_M

(4) **Rotated LBP (RLBP)**

Rotated LBP (RLBP) [20], sometimes calls Dominant Rotated LBP (DRLBP) [21] is rotation technique on LBP around the center pixel. When the reference in the circular neighborhood token by dominant direction, then the weights are assigned with respect to dominant direction. The Fig. is shown the rotation of LBP:

In addition, the RBLP defined by the equation:

$$RLBP_{P,R} = \sum_{p=0}^{P-1} S(g_p - g_c) 2^{mod(p-D,P)} \tag{11}$$

(5) **LBP Histogram Fourier (LBP-HF)**

LBP-HF is is a rotation-invariant image descriptor based on uniform LBPs.

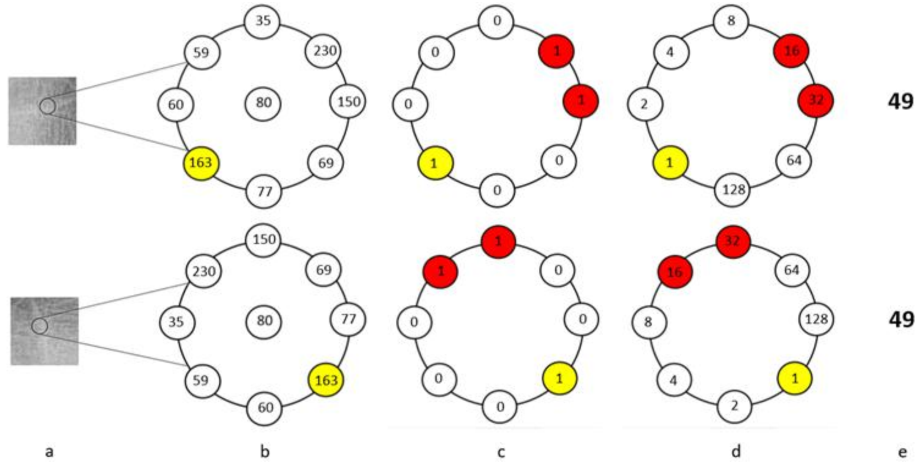

Figure 4: Implementation of RLBP, (a) The 100x100 pixal of knee ROI Texture image (top) and 90° counter-clockwise rotated image (bottom), (b) Yellow color pixel indicate the dominant direction, (c) Values above threshold are shown in red color, (d) The weights are circularly shifted with respect to dominant direction, (e) RLBP values.

The LBP-HF descriptor is formed by first computing a non-invariant LBP histogram over the whole region and then constructing rotationally invariant features from the histogram. LBP-HF is generally used for static features which used Fast Fourier Transform to calculate global features from uniform LBP histogram instated of calculating invariant at each pixel independently. This makes LBP_ri feature set a subset of LBPHF, LBP_ri is discussed in the next sub-section. The Fig.5 is shown the LBP-HF features work:

With the respect to Fig.5,

If =45°, local binary pattern

00000010 $\Rightarrow$ 00000100

00000100 $\Rightarrow$ 00001000,...,

11111000 $\Rightarrow$ 11110001,...,

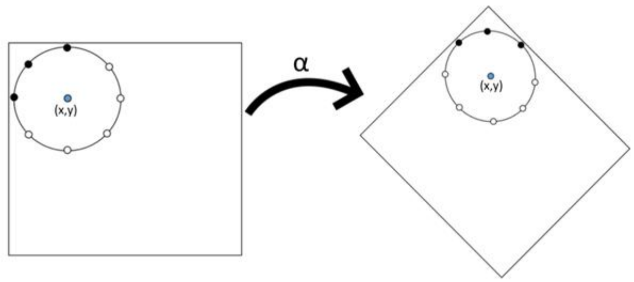

Figure 5: The implementation of LBP-HF feature for rotation invariant image description.

`Similarity if  = K * 45°, as a consequence, the pattern have to be circularly rotated with k steps.`

(6) **LBP Rotation Invariant (LBP_ri)**

LBP_ri is the rotation invariant feature which is based on LBP. The LBP operator produces 2p different output values, corresponding to the 2P different binary patterns that can be formed by the P pixels in the neighbor set. In the research study have applied LBP_ri each pixel with 8 neighbors, LBP_ri with 8 bin can be defined by the equation as :

$$LBP(x,y) = \sum_{P=0}^{7} S(i_P - i_x, y)2^P \tag{12}$$

(7) **Gabor filter**

Gabor filter is another texture extraction technique which used to analysis texture for specific images local region with the specific frequency and specific direction. Gabor filter is in fact of the 2 D Gabor filter bank that consists different parameters including frequencies, orientations and smooth parameters of Gaussian envelope. In addition, Gabor filter bank of pixel (x, y) can be founded by the equation below:

$$G(x,y) \equiv e^{-\frac{(x-x_0)^2}{2\sigma_x^2}-\frac{(y-y_0)^2}{2\sigma_y^2}} e^{j(\omega_{x0}x+\omega_{y0}y)} \tag{13}$$

where

ω_{x0} and ω_{y0} are the centre frequency of x and y direction.

σ_x and σ_y are the standard deviation of the Gaussian function along x and y direction.

(8) **Haralick** Haralick features were determined by using kharalick() function and the basic of Haralick features is the gray-level co-occurrence matrix (GLCM). Haralick features have divided into 14 feature which is calculated from the statistic of basic GLCM including:

- Angular Second Moment (ASM)

$$ASM = \sum_{i=1}^{N}\sum_{j=1}^{N}(P(i,j)^2) \tag{14}$$

- Contrast

$$Contrast = \sum_{n=0}^{N-1} n^2 \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j), |i-j| = n \tag{15}$$

- Correlation

$$Correlation = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(ij)P(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{16}$$

- Variance

$$\sigma^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu)^2 P(i,j) \tag{17}$$

- Inverse Difference Moment(IDM)

$$IDM = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{1+(i-j)^2}P(i,j) \tag{18}$$

- Sum Average

$$SumAverage = \sum_{i=2}^{2N} iP_{x+y}(i) \tag{19}$$

- Sum Variance

$$SumVariance = \sum_{i=2}^{2N}(i-f_8)^2 P_{x+y}(i) \tag{20}$$

- Sum Entropy

$$SumEntropy = -\sum_{i=2}^{2N} P_{x+y}(i) log P_{x+y}(i) \tag{21}$$

- Entropy

$$Entropy = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j) log[P(i,j)] \tag{22}$$

- Difference Variance

$$DifferenceVariance = \sum_{i=0}^{N-1} i^2 P_{x-y}(i) \tag{23}$$

- Difference Entropy

$$DifferenceEnergy = -\sum_{i=0}^{N-1} Px - y(i) log[P_{x-y}(i)] \tag{24}$$

- Information Measure of Correlation 1(IMC1)

$$IMC1 = \frac{HXY - HXY1}{maxHX, HY} \tag{25}$$

- Information Measure of Correlation 2(IMC2)

$$IMC2 = \sqrt{1 - exp[-2(HXY2 - HXY)]} \tag{26}$$

Where

$$\begin{aligned} HXY &= -\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j) log(P(i,j)), \\ HXY1 &= -\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j) log P_x(i) P_y(i), \\ HXY2 &= -\sum_{i=1}^{N}\sum_{j=1}^{N} P_x(i) P_y(j) log P_x(i) P_y(j) \end{aligned} \tag{27}$$

HX and HY are the entropies of P_x and P_y

- Maximum Correlation Coefficient(MCC)

$$MCC = \sqrt{\sum_{k=1}^{N} \frac{P(i,k)P(j,k)}{P_x(i)P_y(j)}} \tag{28}$$

(9) **Local Configuration Pattern (LCP)** Local Configuration Pattern (LCP) is a feature descriptor which separates the architecture of image into two

sates include: 1) local structural information and 2) microscopic configuration (MiC) information that consists of image configuration and pixel-wise interaction relationships. For local structure, information is directly related to the basic of LBP, while the MiC Used to develop for exploring microscopic configuration information. The local structure concept implementation is shown in the fig.6 : With the reference to Figure 6, the Fig.6

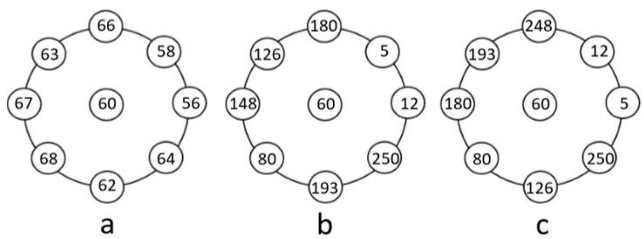

Figure 6: The concept of LCP.

a and Fig.6 b is considered to be the same pattern type by LBP, hence the object's surface would be quite different from each other. On the other hand, with the implementation of LBP with local invariant information; pattern a and b is distinguished, while b and c is considered as the same pattern type due to the same value of variance. In contrast, b and c are different in MiC, which MiC based on textural properties.

For microscopic configuration information is defined as the modeling of microscopic configuration which expresses with the equation:

$$E(a_0, ..., a_{P-1}) = |g_c - \sum_{i=0}^{P-1} a_i g_i| \tag{29}$$

Where

g_c and g_i are intensity center pixel values and neighboring pixels.

a_i(i=0,...,P-1) are weighting parameters associated with g_i

$E(a_0,...,a_{P-1})$ is the reconstruction error regarding model parameters of a_i.

(10) **Local Ternary Pattern (LTP)** With the basic idea of LBP which is analyzed on the central pixel ic which tend to be sensitive to noise particularly

in near-uniform image regions, and to smooth weak illumination gradients, then the idea of LTP is come up to be improved in this issue. LTP can be calculated be gray-level in a zone of width ± t. The s(x) of LBP is replaced by 3-valued function s'(u,i_c,t).

$$s'(u, i_c, t) = \begin{cases} 1 & u \geq i_c + t \\ 0 & |u - i_c| < t \\ -1 & u \leq i_c - t \end{cases} \tag{30}$$

5. Feature Selection and Classification

5.1. Feature Selection

Feature selection is one of the most important in the research work, this process selects the useful feature or reducing data dimensionality to the classification process. There are five different feature selection techniques are applied in the research work:

- Correlation-based Feature Selection (CFS)

  Correlation Feature Selection (CFS) is widely used with the highly correlated to the class feature but CFS produce a low intercorrelation. In addition, in CFS have applied symmetric uncertainty which is the technique use reduce the redundancy of feature. The symmetric uncertainty which applies to two nominal attributes A and B is given by the equation:

$$U(A, B) = 2\frac{H(A) + H(B) - H(A, B)}{H(A) + H(B)} \tag{31}$$

  Where

  H represents as the entropy function.

  H(A, B) the joint entropy of A and B.

  The value of symmetric uncertainty can start from 0 till 1. With the respect to equation (31), CFS can be defined as:

$$CFS = \frac{\sum_{j=i}^{m} U(A_j, C)}{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{m} U(A_i, A_j)}} \tag{32}$$

Where

C refer to the class of feature.

(A_i, A_j) indicates a pair of attributes in the set of features.

- Chi-square ($\tilde{\chi}^2$)

  In the research, the measure of the lacking independence between a feature and a class can be calculated by Chi-square ($\tilde{\chi}^2$). In the implementation of research work by Chi-square ($\tilde{\chi}^2$), the feature is ranked from the most useful to the less. Chi-square ($\tilde{\chi}^2$) can be given as the function below:

$$\tilde{\chi}^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{33}$$

Where

O_{ij} is the observed frequency.

E_{ij} is ij is the expected frequency.

- Information Gain

  In a decision tree, there are two difference node: i) non-terminal nodes refer to test on one or more feature and ii) terminal nodes represented as the output decision. IG use to select the test attribute ate each node. In other words, IG is a feature evaluation method which bases on entropy. IG can be calculated by of a term in the classification of information that can be used. IG can be defined by the equation below:

$$G(D,t) = -\sum_{i=1}^{n} P(C_i) log P(C_i) + P(t) \sum_{i=1}^{n} P(C_i|t) log P(C_i|t) + P(\bar{t}) \sum_{i=1}^{n} P(C_i|\bar{t}) log P(C_i|\bar{t}) \tag{34}$$

Where

C is a set of document collection, feature t. The value of information gain G(D,t) is greater mean t is more useful for the classification for C. This t should be selected.

- Gain Ratio

  Gain ration (GR) is the updating or correction of information gain (IG).

Information gain is used in the decision tree to select the test attribute ate each decision tree node. Hence, the implementation of RG in order to reduce the IGs bias, while choosing an attribute of taking number and size of branches. GR of attribuate(attr) can be written as equation below:

$$GR(attr) = \frac{Gain(attr)}{Entropy(attr)} \tag{35}$$

- ReiefF

  The final feature selection technique in the research refers to relief, relief is a weight based algorithm where the result in the relevant of feature and classification for weight based. In addition, relief will remove the features if the weight is less than the clearly limit. The relief algorithm chooses a sample M in training data S, then the nearest neighbor of the sample H of the sample which has the same group with M, called Near Hit. On the other hand, the reverse categories of M can be found by nearest neighbor sample M from the sample, called Miss.

5.2. Classification

Classification is a process which is used to classify the OA grade, in the processes have applied nine difference machine learning algorithm include:

- Decision Tree(C 4.5)

  Decision tree is a well-known algorithm for machine learning and classification technique. The decision tree has been considered as a direct tree that consists of root node, internal node, and leaf node. For root node is the main root or parent node and has no incoming edges, while leaf root is the root is at the bottom of the tree and has no outgoing edges. For internal node refer to test on an attribute, while branch defines as the output of that test. The Fig.7 shows the decision tree works:

- Average One-Dependence Estimators (AODE)

  Average one-dependence estimators (AODE) is an improvement of the

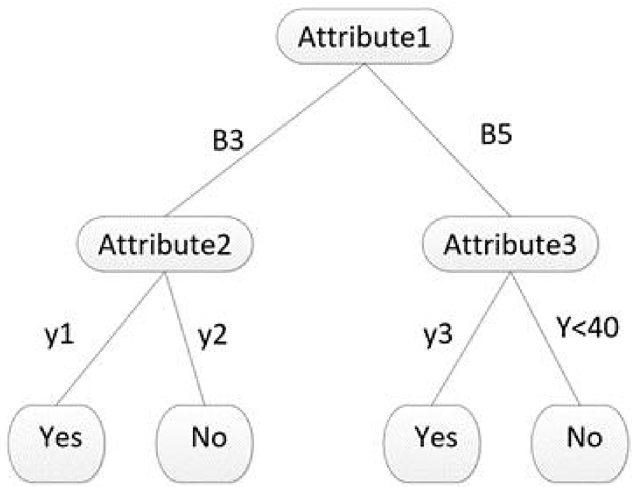

Figure 7: The decision tree.

nave Bayesian classifier and is a probability classification learning technique. AODE comes from the nave Bayes classifier attributed-independence problem. The implement of AODE is to address the attribute-independence problem in nave Bayes. For instance, in the class y which has a set of feature x_1,..., x_n, then AODE can be applied to find the probability of each class y by the equation as follows:

$$\hat{P}(y|x_1, ..., x_n) = \frac{\sum_{i:1\leq i\leq n\wedge F(x_i)\geq m} \hat{P}(y, x_i) \prod_{j=1}^{n} \hat{P}(x_i|y, x_i)}{\sum_{y'\in Y} \sum_{i:1\leq i\leq n\wedge F(x_i)\geq m} \hat{P}(y', x_i) \prod_{i=1}^{n} \hat{P}(x_j|y', x_i)} \tag{36}$$

Where

$\hat{P}()$ is an estimate of P()

F () is the frequency

m is a user specified minimum frequency.

- Bayesian Network (BN)

Bayesian network (BN) is a graphical probability model used for reasoning and the decision making in uncertainty. In other words, the Bayesian network is a directed acyclic graph (DAG) and each node n ∈ N of BN represents a domain variable or dataset attribute. In addition, the Bayesian network highly depends on Bayes rule. The Bayes' rule can be written as follows: Assume A_i attribute where i= 1,...,n, and take value a_i where i= 1,...,n

Assume C as class label attribute and U=(a_1,..., a_n) as unclassified test instance. U will be classified into class C based on Bayes rule is represented as:

$$P(C|U) = arg\ maxP(C)P(U|C) \tag{37}$$

- Nave Bayesian Classifier

  Nave Bayesian is one of the well-known of Bayesian techniques which called state-of-the-art of the Bayesian. In the work of nave Bayesian classifier have assumed all the attribute in the same class has been considered as independent given class label. With the respect to Bayes rule, the Nave Bayesian has been modified as the equation below:

$$P(C|U) = arg\ maxP(C)\prod_{i=1}^{n} P(A_i|C) \tag{38}$$

- Support Vector Machine(SVM)

  Support vector machine (SVM) is a popular liner classifier and widely used for the classification task. SVM is performed for the best separating by constructing an N-dimensional hyperplane between two training sample classes in the feature set. In SVM classifier have divided into two categories:

  - Linear classification

    In the liner classification, the SVM can be divided into two type of classification: i) linear separable case and ii) linear non-separable case. In linear separable case, SVM with the training data x_i , y_i, y_i ∈-1,+1 , i = 1,...,n , can be defined as the equation as follows:

$$\begin{cases} x_i.w + b \geq +1 & \text{for} y_i = +1 \\ x_i.w + b \leq -1 & \text{for} y_i = -1 \end{cases} \tag{39}$$

For the linear non-separable case, SVM equation is changed to:

$$\begin{cases} x_i.w + b \geq +1 - \xi_i & \text{for} y_i = +1 \\ x_i.w + b \leq -1 + \xi_i & \text{for} y_i = -1 \\ \quad \xi_i \geq 0, i = 1, n \end{cases} \tag{40}$$

– Nolinear classification

In nolinear classification, svm equation can be written as:

$$f(x) = \sum_{i=1}^{n_s} \alpha_i y_i P(x_i, x) + b \tag{41}$$

Where

n_s is the number of support vector.

is non-negative Lagrange multipliers

P (x, y) is Polynominal of degree m: $k(x,y)=(x.y+1)^m$

- Logistic Regression

  Logistic regression is a well-known statistic regression model and offshoot the ordinary regression. in the research, the logistic regression has been applied to the dependent variable. In addition, in logistic regression used logistic function, also recognized as the sigmoid function which use to compute logistic model. The logistic model has defined with the equation:

$$logit(p_t) = x_t b + e_t; p_t = P(y_t = 1|x_t, b) \tag{42}$$

  Where

  yt 0,1 x_t is the regression vector, x_t = [1, x_1,x_2,..., x_n]

  b is the model parameters vector, b = [b_0, b_1,..., b_n] P(—) is the conditional probability function (pf) and logit() is the logistic function. The logit() can be represented as:

$$logit(p) = ln(\frac{p}{1-p}) \tag{43}$$

- The Back Propagation Algorithm

  Sequential Minimal Optimization (SMO) is the improvement from SVM to solve the quadratic programming (QP) optimization problem, the problem has happened during the SVM training. Furthermore, QP problem can be represented while SVM finds α_i , as the equation below:

$$min\ L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i x_j) \tag{44}$$

$$s.t\ 0 \leq \alpha_i \leq C \quad and \quad \sum_{i}^{n} \alpha_i y_i = 0, \quad \forall i \tag{45}$$

With the respect to the two equation above (44) and (45), SMO select the Smallest Optimization Problem (SOP) and solve the SOP.

- The Back Propagation Algorithm

  The backpropagation algorithm is a famous in the neural network, the implementation of backpropagation algorithm is to calculate a gradient which is demanded in the calculation of the weights to be used in the network. In other words, backpropagation is doing the task of learning the same to the neural network which is finding the minimum of the error function.

6. Data Collection

6.1. Dataset

The dataset of the research is the research image which is collected from 2 different hospitals include: Dibuk hospital and Bangkok Hospital where located in Phuket province, Thailand. The number of images in the research use 131 images which divided into non-OA 63 images and OA has 68 images. Due to the privacy of each patient for each image, the researcher request only the image data without including any detail information for example age, sex, address, and etc.

6.2. Dataset Classification and Region of Interest

The dataset is collected with the known result of OA and non-OA case, which the OA case has 68 images in 131 images of all dataset. Furthermore, four different places are chosen to be ROI for texture analysis, the four ROIs are shown in the Fig.8 below:

With the reference to the Fig. which has 4 different ROI, then the dataset of

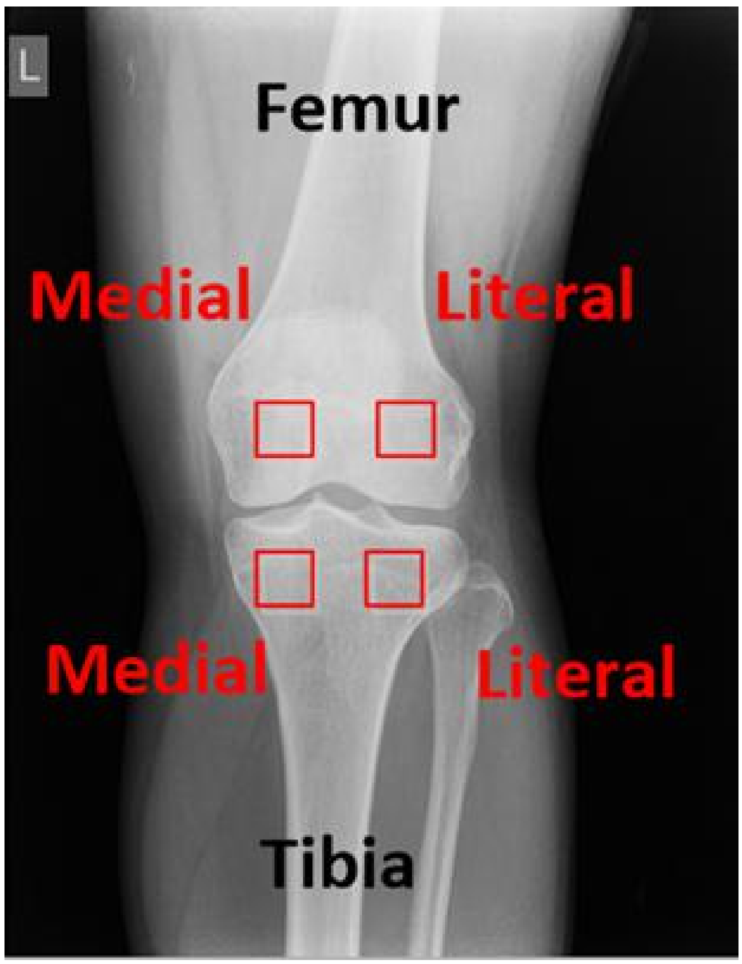

Figure 8: The Four ROI of Texture Analysis

the research is divided into four different datasets which have 131 ROI images include i)Medial Femur ROI dataset, ii) Literal Femur ROI dataset, iii) Medial

Tibia dataset, and iv) Literal Tibia dataset.

The four ROI datasets which are used to analyze and evaluate with image processing techniques and classification techniques present in section VII) Evaluation.

7. Evaluation

In this section, the experiment setup and objective of evaluation in result experiment is presented. The experiment setup in this section is mention all the number of experiments. The number of the experiment is calculated by:

ROI has four 4 regions

Feature selection: 5 algorithms and the value of K=10, 20, 30, 40, 50, and 60;

Feature Descriptor: 10 descriptors

Learning algorithm has 9 algorithms

With the respect to feature descriptor the number of experiment is shown as follows:

CLBP, Gabor, LBP, LBP-hf, RLBP, LCP, and LTP have 900 experiments for each

LBP_ri has 612 experiments

GLCM and Haralick have 180 experiments for each.

Hence, all the number of research experiments = (900x7) + 612 + (180x2)= 7272 experiments.

The evaluation of each experiment is measured by six measurable parameters include Area Under Curve (AUC), Accuracy (AC), Sensitivity (SN), Specificity (SP), Precision (PR), and F-Measure.

The objective of the research experiment is discussed as followed.

7.1. The best ROI result

A four different places of ROI for instance: Literal Femur(LF), Medial Femur(MF), Literal Tibia(LT), and Medial Tibia(MT) are selected to use for texture analysis to define the OA or non-OA case.The result of each ROI is shown in table 1.

Algorithm / ROI	AUC	AC	SN	SP	PR	FM
Literal Femur(LF)	0.912	0.832	0.832	0.832	0.832	0.832
Medial Femur(MF)	0.884	0.794	0.794	0.792	0.794	0.794
Literal Tibia(LT)	0.883	0.809	0.809	0.809	0.809	0.809
Medial Tibia(MT)	0.895	0.802	0.802	0.802	0.802	0.802

Table 1: The ROI best result comparison.

With the reference to table 1, the best result of LF is founded by LBP feature descriptor apply with Bayes Network. On the other hand, the best result of MF is pointed out by RLBP apply with Nave Bay, while the best result of LT is given by Gabor applied with Bayes Network for AUC values and other highest values are founded by LBP and RLBP applied with Bayes Network. Finally, the best result of MT is given by RLBP applied with Bayes Network. The best feature descriptor is presented in the next subsection.

7.2. The best feature descriptor result

In this subsection, there are ten of texture descriptor is implemented in the research include, i) Histogram features, ii) Local Binary Pattern (LBP), iii) Completed LBP (CLBP), iv) Rotated LBP (RLBP), v) LBP Histogram Fourier (LBP-HF), vi) LBP Rotation Invariant (LBP_ri), vii) Gabor, viii) Haralick, ix) Local Configuration Pattern (LCP) and x) Local Ternary Pattern (LTP). The result of each feature descriptor is chosen the best result in the research experiment result. The further detail is pointed out Table.2 below:

With the respect to Table2. is selected for the best result of four different ROI. For feature selection, the CFS produce the best result in the feature selection techniques. First and foremost, the result of First Level of GLCM is

Texture Descriptor \ Algorithm	AUC	AC	SN	SP	PR	FM
Histogram	0.757	0.695	0.695	0.69	0.695	0.693
CLBP	0.882	0.763	0.763	0.762	0.763	0.763
Gabor	0.883	0.786	0.786	0.786	0.786	0.786
Haralick	0.695	0.664	0.664	0.67	0.672	0.662
LBP	0.912	0.832	0.832	0.832	0.832	0.832
LBP-hf	0.773	0.71	0.71	0.717	0.71	0.709
LBP_ri	0.812	0.771	0.771	0.771	0.771	0.771
LCP	0.783	0.725	0.725	0.724	0.725	0.725
LTP	0.816	0.756	0.756	0.761	0.763	0.755
RLBP	0.895	0.809	0.809	0.81	0.81	0.809

Table 2: The ROI best result comparison.

pointed out by LF ROI have applied with Bayes Network algorithm and the result of CLB is presented from two different ROI, the highest value of AUC produce by LF applied with Bayes Network algorithm while others value are produced by MT applied with Bayes Network. In addition, the best result of Gabor is given by LT applied with Bayes and MT is applied to Bayes Network for others. Furthermore, the Haralick best result is selected from LF apply with Bayes Network for AUC value and MF apply with AODE for others. The selection from LF have applied with Bayes Network can produce the best result of LBP. Moreover, the best result of LBP-hf is produced by MF applied with AODE, while the LBP_ri best result is pointed out by LF with Bayes Network algorithm for AUC value and others value by LF with Multilayer algorithm. For LCP best result, is given by MT with Nave bay algorithm for AUC value and others from LT applied with decision tree with binary true. In LTP best result, the result is founded by LF applied with Bayes Network for AUC value, while the AC, SN, and FM get from LT and LF are applied with Bayes Network, and lastly, SP and PS value is given by LF applied with Bayes Network. Finally, RLBP best result is presented by MT applied with Bayes Network for AUC value, PR value is pointed out by LT have applied with Nave bay, and others are given by LT applied with Bayes Network and Nave Bay.

7.3. The Best feature selection

Back to feature selection process, there is five feature selection techniques are presented in the work, for instance: i) Correlation-based Feature Selection (CFS), ii) Chi-square, iii) Gain Ration, iv) Information Gain and v) Relief. In addition, the result of each algorithm include i) Chi-square, ii) Gain Ration, iii) Information Gain and iv) Relief remains the same result of the different K value. This sub-section is presented more detail in Table. 3:

Algorithm / Texture Selector	AUC	AC	SN	SP	PR	FM
CFS	0.912	0.832	0.832	0.832	0.832	0.832
Chi-Square	0.699	0.687	0.687	0.687	0.687	0.687
Gain Ratio	0.709	0.687	0.687	0.687	0.687	0.687
Information Gain	0.699	0.687	0.687	0.684	0.687	0.687
Relief	0.699	0.679	0.679	0.674	0.681	0.677

Table 3: The best feature selection result comparison.

With reference to table 3, the best result of CFS can found out by LF applied with LBP and Bayes Network, while Chi-square result is presented by LF applied with GLCM and Nave Bay for AUC value and others value from MT applied with CLBP and decision tree. For the gain ratio, the best result, have founded by LT applied with LBP_ri and decision tree binary true, and others value are pointed out by LT applied with LTP and SMO. In the information gain, the best result is presented by LF applied with GLCM and Nave Bay, while others value by LT applied with LTP and SMO. Lastly, the best result of Relief is received from LF applied with GLCM and Nave Bay. The best result of the learning algorithm is presented in next subsection.

7.4. The Best learning algorithm

The learning algorithm is used in the study include: i) Decision Tree(C4.5), ii) Decision with binary true, iii) Average One-Dependence Estimators (AODE), iv) Bayes Network, v) Nave Bay, vi) Support Vector Machine (SVM), vii) Logistic, viii) Sequential Minimal Optimization, and ix) backpropagation algorithm.The best result of each algorithm is shown in Table. 4:

Algorithm / Texture Descriptor	AUC	AC	SN	SP	PR	FM
C4.5	0.757	0.779	0.779	0.78	0.78	0.779
C4.5 binary tree(true)	0.766	0.74	0.74	0.736	0.742	0.739
AODE	0.896	0.809	0.809	0.804	0.809	0.809
Bayes network	0.912	0.832	0.832	0.832	0.832	0.832
Nave bay	0.903	0.817	0.817	0.816	0.817	0.817
SVM	0.715	0.718	0.718	0.711	0.72	0.715
Logistic	0.904	0.84	0.84	0.844	0.847	0.839
SMO	0.771	0.771	0.771	0.771	0.771	0.771
Multilayer	0.851	0.771	0.771	0.77	0.771	0.771

Table 4: The best learning algorithm result comparison.

With the reference to Table 4, the best result of each learning algorithm is introduced. In the best result of decision tree have gained by the implementation of MT applied with CLBP and CFS, while decision tree with the binary true best result is pointed out by LT applied with LBP and CFS for AUC value and others from LTP applied with LTP and CFS. For the AODE result is presented by LF applied with LBP and CFS for AUC value, LT applied with LBP and CFS for SP value, and others are founded by LT or LF applied with LBP and CFS. The implementation of LF applied with LBP and CFS have produced the best result of Bayes Network, while the best result of Nave bay is produced by LF applied with LBP and CFS for AUC value and others values by LT applied with LBP and CFS. Furthermore, the best result of SVM can be founded out by the implementation of LF applied with LBP_ri and CFS, while the logistic result is presented by the implementation of LF applied with LBP and CFS. For the SMO best result is a reference to the implementation of LF applied with RLBP and CFS. Lastly, the multilayer finds out by LF applied with RLBP and CFS for AUC, LT and others are applied from LF applied with LBP_ri with CFS.

8. Conclusion

OA has been spreading in elder people society around the world. The best way to prevent of OA is to check up for early detection. X-ray is the simplest solution to detect OA due to the cost and performance x-ray imaging. The implementation and classification techniques to x-ray image for OA detection is considered as one of the challenging issues in IT research filed. With the purpose of the research are divided into 4 purpose: i) finding the best ROI of bone texture which produces the best result of OA detection, ii) the best texture descriptor amount of nine in the research study, iii) The Best feature selection in the research and iv) The Best learning algorithm for the research study. Finally, the result of the research purpose show that the ROI of literal femur(LF) is the result of the first purpose, LBP solves for texture descriptor, CFS is selected as the best feature selection in the research study, and Bayes Network performance the best in classification for research.

9. Acknowledgement

We would like to our gratitude to the Bangkok hospital Phuket branch and Dibuk hospital for providing x-ray dataset in this research study. We would like to give a special thanks to MD. Sirisak Yaisoongnern for managing to the process of data collection and million thanks to MD. Chaowakon Saehang for sharing his valuable knowledge in OA grading field.

References

[1] I. R. Factors, NIH Conference Osteoarthritis : New Insights AND I TS, Annals of Internal Medicine 133 (8) (2000) 637–639. doi:10.7326/0003-4819-133-8-200010170-00016.
URL http://www.annals.org/content/133/8/635.short

[2] Bone and Joint Initiative USA, The Burden of Musculosketal Diseases in the United States: Prevalence, Societal and Economic Cost, 2014.

[3] M. Wolski, P. Podsiadlo, G. W. Stachowiak, L. S. Lohmander, M. Englund, Differences in trabecular bone texture between knees with and without radiographic osteoarthritis detected by directional fractal signature method, Osteoarthritis and Cartilage 18 (5) (2010) 684–690. doi:10.1016/j.joca.2010.01.002.

[4] Z.-J. X. K.-M. L. J. L. Chao Jin, Yang Yang, Automated analysis method for screening knee osteoarthritis using medical infrared thermography chao, Journal of Medical and Biological Engineering 33 (5) (2013) 471. doi:10.5405/jmbe.1054.
URL http://jmbe.bme.ncku.edu.tw/index.php/bme/article/view/1962/1005

[5] L. Shamir, S. M. Ling, W. Scott, M. Hochberg, L. Ferrucci, I. G. Goldberg, Early detection of radiographic knee osteoarthritis using computer-aided analysis, Osteoarthritis and Cartilage 17 (10) (2009) 1307–1312. arXiv:NIHMS150003, doi:10.1016/j.joca.2009.04.010.
URL http://dx.doi.org/10.1016/j.joca.2009.04.010

[6] S. S., P. U., R. R., Detection of Osteoarthritis using Knee X-Ray Image Analyses: A Machine Vision based Approach, International Journal of Computer Applications 145 (1) (2016) 20–26.

doi:10.5120/ijca2016910544.
URL http://www.ijcaonline.org/archives/volume145/number1/gornale-2016-ijca-910544.pdf

[7] M. Kotti, L. D. Duffell, A. A. Faisal, A. H. McGregor, Detecting knee osteoarthritis and its discriminating parameters using random forests, Medical Engineering and Physics 43 (2017) 19–29. doi:10.1016/j.medengphy.2017.02.004.
URL http://dx.doi.org/10.1016/j.medengphy.2017.02.004

[8] G. Castellano, L. Bonilha, L. M. Li, F. Cendes, Texture analysis of medical images, Clinical Radiology 59 (12) (2004) 1061–1069. doi:10.1016/j.crad.2004.07.008.

[9] T. Janvier, H. Toumi, K. Harrar, E. Lespessailles, R. Jennane, ROI impact on the characterization of knee osteoarthritis using fractal analysis, 5th International Conference on Image Processing, Theory, Tools and Applications 2015, IPTA 2015 (2015) 304–308doi:10.1109/IPTA.2015.7367152.

[10] T. K. Chuah, C. L. Poh, K. Sheah, Quantitative texture analysis of MRI images for detection of cartilage-related bone marrow edema, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (2011) 5112–5115doi:10.1109/IEMBS.2011.6091266.

[11] K. Dittakan, F. Coenen, Early detection of osteoarthritis using local binary patterns: A study directed at human joint imagery, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9810 LNCS (2016) 93–105. doi:10.1007/978-3-319-42911-3_8.

[12] N. N. Kachouie, P. Fieguth, A medical texture local binary pattern for TRUS prostate segmentation, Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings (2007) 5605–5608doi:10.1109/IEMBS.2007.4353617.

[13] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classificatio, IEEE Transactions on Image Processing 19 (6) (2010) 1657–1663. doi:10.1109/TIP.2010.2044957.

[14] N. M. Varney, V. K. Asari, Volumetrie features for object region classification in 3D LiDAR point clouds, Proceedings - Applied Imagery Pattern Recognition Workshop 2015-February (February). doi:10.1109/AIPR.2014.7041941.

[15] P. S. Prasad, B. P. Rao, LBP-HF features and machine learning applied for automated monitoring of insulators for overhead power distribution lines, 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (2016) 808–812doi:10.1109/WiSPNET.2016.7566245.
URL http://ieeexplore.ieee.org/document/7566245/

[16] X. Tan, B. Triggs, Recognition Under Difficult Lighting Conditions, IEEE transactions on image processing 19 (6) (2010) 1635–1650. doi:10.1109/TIP.2010.2042645.
URL http://www.ncbi.nlm.nih.gov/pubmed/20172829

[17] Z. Wang, R. Huang, W. Yang, C. Sun, An enhanced Local Ternary Patterns method for face recognition, Proceedings of the 33rd Chinese Control Conference, CCC 2014 0 (1) (2014) 4636–4640. doi:10.1109/ChiCC.2014.6895720.

[18] P. P. Kawathekar, Use of textural and statistical features for analyzing severity of radio-graphic osteoarthritis of knee joint, 2015, pp. 1—4.

[19] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1) (1996) 51–59. doi:10.1016/0031-3203(95)00067-4.

[20] R. Mehta, K. Egiazarian, Rotated Local Binary Pattern (RLBP) - Rotation Invariant Texture Descriptor., Icpram (2013) 497–502.

URL http://dblp.uni-trier.de/db/conf/icpram/icpram2013.html{#}MehtaE13

[21] R. Mehta, K. Egiazarian, Dominant Rotated Local Binary Patterns (DRLBP) for texture classification, Pattern Recognition Letters 71 (2016) 16–22. doi:10.1016/j.patrec.2015.11.019.
URL http://dx.doi.org/10.1016/j.patrec.2015.11.019