Image Texture Analysis for Medical Image Mining: A Comparative Study Direct to Osteoarthritis Classification using Knee X-ray Image

Sophal Chan[a,1], Kwankamon Dittakan[a,2], Matias Garcia-Constantino[b]

[a] College of Computing, Prince of Songkla University, Phuket Campus Kathu, Phuket, 83120, Thailand
E-mail: [1]sophal.c@phuket.psu.ac.th; [2]kwankamon.d@phuket.psu.ac.th

[b] School of Computing, Ulster University, BT37 0QB, United Kingdom
E-mail: m.garcia-constantino@ulster.ac.uk

Abstract— Knee Osteoarthritis (OA) is one of the most prominent diseases in an ageing society and has affected over 10 million people in Thailand. When people suffer from OA, it is very difficult to recover. Therefore, early detection and prevention are important. The typical way to detect OA is by using X-ray imaging. This research study is focused on early detection of OA by applying image processing and classification techniques to knee X-ray imagery. The fundamental concept is to find a region of interest, use feature extraction techniques and build a classifier that can classify between OA or non-OA imageries. There are four regions of interest obtained from each image: (i) Medial Femur (MF), (ii) Lateral Femur (LF), (iii) Medial Tibia (MT), and (iv) Lateral Tibia (LT). The ten texture analysis techniques are then adopted to generate the embedded properties of the bone surface. Once the feature vector has been generated the variety of techniques of machine learning mechanisms are applied to generate the desired classifiers, which can be used to distinguish between OA and non-OA images. From the conducted experiments, a total of 131 images (68 OA cases and 63 non-OA cases) was used, the results obtained show that LF region with Local Binary Pattern descriptor produced the most appropriate classifier with an AUC value of 0.912.

Keywords— texture analysis; osteoarthritis; knee OA; image classification.

I. INTRODUCTION

Osteoarthritis (OA) is considered a degenerative disease of human joints. OA is the most prevalent disease of the joints in the ageing society and the most typical disease of arthritis which affects millions of people in the United States[1]. In Thailand, the people affected with OA has almost reached 10 million, which is 13% of Thailand's population, as reported by the National Statistical Office (NSO) in 2014. The NSO also reports that Thailand is going to be an ageing society country in ten years. In the ageing society, knee OA affected approximately 10% of men and 13% of women in 2010 [2]. Symptoms of knee OA can be detected by the presence of pain, swelling, stiffness in the knee, reduced ability of movement and cracking sound when the knee is moved. Furthermore, OA can be early detected using medical images to prevent the progression to a more severe stage. Medical imaging that is widely used for OA early detection includes (i) X-ray image, (ii) Computed Tomography (CT) and (iii) Magnetic Resonance Imaging (MRI). In this work, the X-ray image is suggested due to the widely used in Thailand. Figure 1 below shows knee X-ray images for the (a) Normal and (b) OA case.

Fig. 1 Normal and Osteoarthritis (OA) Knee X-ray Imagery

The fundamental idea of this research is to apply image processing techniques on medical X-ray images to detect OA. Image processing is a technology in which algorithms can be used to enhance the image or extract some useful features

from the image to study for any specific purpose. Image processing is typically used in many researches, such as biology, medicine, astronomy, and biometrics. In other words, image processing is a specific technology used for: (i) classification, (ii) feature extraction, (iii) multi-scale signal analysis, (iv) pattern recognition and (v) projection. The implementation of image processing for classification in medical X-ray images is discussed in this work.

The objective of the research is mainly focused on the classification of the OA and non-OA X-ray images from four sub-images of the knee in terms of the following regions of interest (RoI): (i) Lateral Femur (LM), (ii) Lateral Tibia (LT), (iii) Medial Femur (MF), and (iv) Medial Tibia (MT). The image can be analyzed in three main different ways (i) color-based analysis: image can analyze based on the color in the image, (ii) shape analysis: image can analyze based on the shape of the object in the image, and (iii) texture analysis: image can analyze based on the structure information of the object in the image. Texture analysis is one of the appropriate ways to analyze X-ray images because the X-ray image is the binary image (black/white color image). Texel is the basic unit of a graphic in terms of texture. A texture is a set of Texel's occurring in some regular or repeated pattern. In other words, texture can produce information about the spatial arrangement of intensity in an image or selected image regions. In addition, texture can be analyzed by texture descriptors, which is a technique that can use statistics, filter banks, auto-correlation, etc. There are ten texture descriptors were applied in the work: (i) histogram feature, (ii) Local Binary Pattern (LBP), (iii) Completed LBP (CLBP), (iv) Rotated LBP (RLBP), (v) LBP Rotation Invariant (LBP$_{ri}$), (vi) LBP Histogram Fourier (LBP-HF), (vii) Local Ternary Pattern (LTP), (viii) Local Configuration Pattern (LCP), (ix) Haralick feature, and (x) Gabor filter feature. The detail is given in Section 5.

The proposed mechanism offers three major advantages: (i) fast speed processing, (ii) can be applied in a future study with different imaging modalities, and (iii) can help non-specialist researchers or new profession MD (Medical Doctor) to analyze OA/non-OA images. On the other hand, the proposed framework still has two main limitations: (i) it is a semi-automatic system which needs an action from a human to identify the RoI and (ii) the experiment applied with the small dataset.

The remainder of this paper is arranged as follows. The materials and methods of the research are presented in Section 2. Section 3 describes the result and discussion of the works. Enduringly, the study conclusion is described in Section 4.

II. MATERIALS AND METHOD

In this section, the materials and proposed methods of the research are presented. In this section is separated into two different sub-sections: Sub-section A presents the related work, while the proposed framework is described in Sub-section B. The dataset used in this research and the process of identification of the region of interest are mentioned in Sub-section C. In the context of this research texture analysis mechanism is suggested to analyze the knee X-ray image. Thus, the description of texture analysis is presented in Sub-section D. The brief explanation of the algorithms used in

dimension reduction process is discussed in Sub-section E. Finally, the classification techniques which is used to generate the desired classifiers are presented in Sub-section F.

A. Related Works

In recent years there has been substantial research work in OA detection and classification [3]–[7]. The early detection of OA can be applied to medical imaging and used in conjunction with a professional clinician to classify the OA. On the other hand, the implementation of classification techniques to medical imaging for OA detection has been considered as an interesting topic in image processing research. Texture plays as one of the most important properties in images as it shows the arrangement of pixels in objects to analyze. Therefore, the texture is one of the useful solutions in medical image processing for diagnosis and detection of OA in a clinical setting[6], [7]. In [4], the research focused on the RoI of tibia texture for the analysis of OA. Texture analysis can be applied to various types of medical images, including X-ray [4], MRI [8], CT[9], and Infrared[10]. In addition, the texture of an object can be analyzed using texture descriptors, which is a technique to represent and handle texture in a numeric form. Texture descriptors include: LBP [7], [11]; CLBP[12]; LBP$_{ri}$; LBP-HF[13] and LTP[14], [15].

The study presented in [3] introduced a method to analyze knee OA X-ray images which combined different types of features: (i) shape, (ii) statistical, (iii) Haralick, (iv) texture analysis and (v) first-four moments features. The classification algorithm used in [3] was Random Forest, with the data divided into 40% for training and 60% for testing. The features considered in [16] to analyze texture for radiographic OA of knee joint were: (i) entropy, (ii) mean, (iii) median, (iv) standard deviation, (v) variance, and (vi) Tamura texture features.

Furthermore, texture analysis applications have been widely applied in the medical research for the detection of various diseases including: tumor heterogeneity [9], [17], [18]; brain tumor [19], [20]; head and neck cancer[21], [22]; emphysema[23], [24]; prostate segmentation [25]–[27], colon cancer[28], [29]; small vessel disease and blood brain barrier [30], breast cancer [31]–[34]; skin cancer [35]–[37] retinal vessel segmentation[38], [39] and lung cancer [40], [41].

B. Proposed Framework

The proposed framework is presented in Figure 2. The figure shows the three main processes considered: (i) Region of interest (RoI) identification, (b) texture extraction, and (c) classification. First and foremost, in order to detect OA/non-OA imagery by using texture analysis, it is required to have good quality sub-images of texture. RoI selection is used to select a specific area or sub-image, which is considered to have a unique identity for detecting OA by using texture analysis. The RoIs in this process are selected from four different regions as shown in Figure 2 (a): (i) two on the femur bone on the lateral and medial side, and (ii) two on the tibia on the lateral and medial side. The output of the RoI selection process is the four RoIs: (i) Medial Femur (MF) RoI, (ii) Lateral Femur (LF) RoI, (iii) Medial Tibia (MT)

RoI, and (iv) Lateral Tibia (LT) RoI. The output of the RoI selection process was used as the input of the texture extraction process.

Texture extraction (refers to Figure 2(b)) was used to extract the texture of each sub-image from Figure 2(a). Feature extraction is a process used to extract the feature by using texture descriptor and decrease the size of feature space from extracting in terms of the number of values and dimensions by applying feature selection to produce the feature vector. Feature descriptor is a technique widely used to extract features.
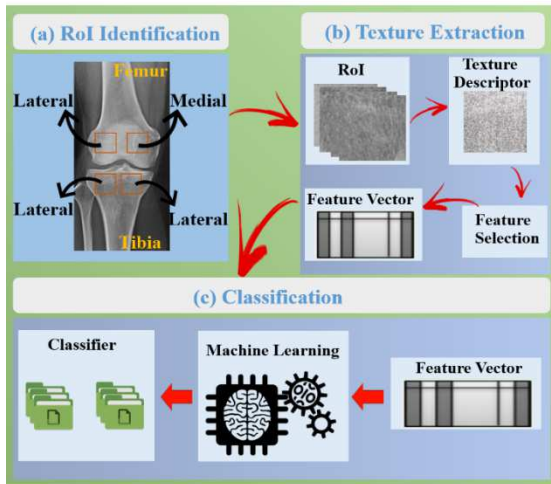

Fig. 2 Texture Analysis on Knee OA Classification Techniques

The research study presented here considered ten different feature extraction techniques, with seven of them belonging to the Local Binary Pattern (LBP) family. LBP is a popular texture descriptor technique which is applied to analyze the center pixel in relation to neighboring pixels. In addition to LBP, another six feature descriptors related to LBP applied in this research are: (i) Completed LBP (CLBP), (ii) Rotated LBP (RLBP), (iii) LBP Rotation Invariant (LBP_{ri}), (iv) LBP Histogram Fourier (LBP-HF), (v) Local Ternary Pattern (LTP) and (vi) Local Configuration Pattern (LCP). Besides the texture descriptors from the LBP family, the authors implemented three other descriptors: (i) histogram feature, (ii) Haralick feature, and (iii) Gabor filter feature. The implementation of the ten feature descriptors was implemented in MATLAB. When texture descriptor was applied, the feature space is produced. Feature selection was applied in order to reduce the feature space in term of number and dimensionality. The output of this process is a feature vector which is used as the input of the classification process.

Classification is the final process of the proposed framework (refers to Figure 2(c)). In order to classify OA/non-OA images, it is required to use machine learning algorithms to apply with the feature vector that was generated in the texture extraction process. Nine machine learning methods were applied: (i) Decision Tree (C4.5), (ii) Binary Split Tree, (iii) Average One-Dependence Estimators (AODE), (iv) Bayesian Network (BN), (v) Naïve Bayes Classier, (vi) Support Vector Machine (SVM), (vii) Logistic Regression, (viii) Sequential Minimal Optimization (SMO), and (ix) Backpropagation. The implementation of the

machine learning algorithms on the feature vector was then evaluated using several typically used evaluation measures.

C. Dataset and Region of Interest (RoI) identification

In this section, the knee X-ray dataset collection, and the analysis of the region of interest (RoI) of the knee for classification are presented.

1) Dataset: The dataset applied in this research study was collected from two local hospitals in Thailand. The number of images used was 131, and they were divided as follows: (i) 63 non-OA images, and (ii) 68 OA images. To protect the data privacy of the patients that participated in this study, for each image collected only the image data was used, thus excluding personal information (e.g. age, gender, address, etc.).

2) Region of Interest (RoI): Four different places in the knee X-ray images were identified and used as RoI for texture analysis, and they are shown in Figure 3:
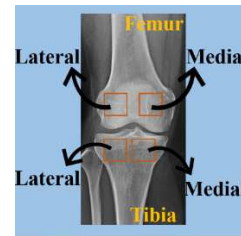

Fig. 3 The Four RoI of Texture Analysis

Once four different RoIs were identified from a collection of X-ray images. Therefore, four image data sets were generated. Each set containing the 131 images and could be named as (i) Medial Femur (MF) dataset, (ii) Lateral Femur (LF) dataset, (iii) Medial Tibia (MT) dataset, and (iv) Lateral Tibia (LT) dataset. The four RoI datasets were used for analysis and evaluation using image processing and classification techniques.

D. Texture Analysis

A brief detail of ten texture descriptors used with reference to the work illustrated in this paper is presented in this section. Texture descriptor is a technique to characterize image textures or regions. Texture descriptor can observe the region of interest in images or specific the region border. Texture feature is one of the most important image features which is used for image mining and image classification. The texture is a feature that gives the information of intensities in an image. In this section, ten texture feature descriptors are discussed for extracting texture feature: (i) histogram feature, (ii) Local Binary Pattern, (iii) Completed LBP, (iv) Rotated Local Binary Pattern, (v) Local Binary Pattern Rotation Invariant, (vi) Local Binary Pattern Histogram Fourier, (vii) Local Ternary Pattern, (viii) Local Configuration Pattern, (ix) Haralick feature, and (x) Gabor filter feature descriptor. Each texture descriptor is presented in the following sub-section.

1) Histogram Feature: The histogram feature of the grey level image is defined by the state-of-the-art histogram-based feature, the histogram feature used in work, including (i) Mean, (ii) Variance, (iii) Skewness, (iv) Kurtosis, (v)

Energy, (vi) Entropy. Each feature can be defined using the equation (1) to (7) respectively.

- Mean: is the average of feature i in the grey level of the image. Mean can be defined as the following equation

$$\mu = \sum_{i-1}^{n} iP(i) \qquad (1)$$

Where $P(i)$ refers to the probability distribution of bin i, which $P(i)$ can be defined as:

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

- M *is the block number, and* H(i) *is the histogram function*

- Variance*: refers to the measuring of the histogram width that measures the deviation of grey levels from the Mean

$$\sigma^2 = \sum_{i=1}^{n} (i - \mu)^2 \, P(i) \qquad (3)$$

where σ^2 is the variance and μ is the mean.

- *Skewness:* is used for measuring the degree of histogram asymmetry around the Mean

$$skew = \frac{1}{\sigma^3} \sum_{i=1}^{n} (i - \mu)^3 \, P(i) \qquad (4)$$

where σ refers to the standard deviation, which is the square root of the variance presented in equation (3).

- *Kurtosis:* refers to the histogram sharpness measuring

$$Kurtosis = \frac{1}{\sigma^4} \sum_{i=1}^{n} (i - \mu)^4 \, P(i) \qquad (5)$$

- *Energy:* is applied for describing the estimation of information in an image

$$Energy = \sum_{i=1}^{n} [P(i)]^2 \qquad (6)$$

- *Entropy:* is used for randomness measure and takes low values for smooth images

$$Entropy = -\sum_{i=1}^{n} P(i) \, log_2[\, P(i)] \qquad (7)$$

2) Local Binary Pattern: Local Binary Pattern (LBP) [42] is applied to label the pixel. LBP operator compares the intensity value of the center pixel with the surrounding neighborhoods. The output is the binary number. The basic LBP is operated illustrated in Figure 4 below.
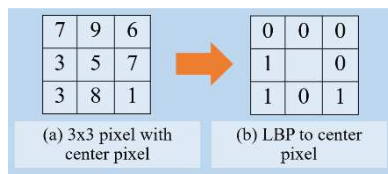

Fig. 4 LBP Operator

In addition, LBP at pixel (x$_c$, y$_c$) can be calculated by the equation below:

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} S(i_p - i_c) \rightleftarrows \rightleftarrows \rightleftarrows \rightleftarrows 2^P \qquad (8)$$

where: *P* is the pixels, *R* is a radius of the circle, i_c and i_p are the grave-level values of the center point in the pixel *P*, *S(x)* is a function which is represented following equation.

$$S(x) = \begin{cases} 1 & if \rightleftarrows x \geq 0 \rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows \\ 0 & if \rightleftarrows x < 0 \end{cases} \qquad (9)$$

3) Completed Local Binary Pattern: In CLBP, a local region is defined by a center pixel with a local difference sign-magnitude transform (LDSMT) [12]. This research study is focused on LDSMT, which breaks down the local structure of images in two elements: (i) the difference signs (CLBP_S), and (ii) the difference magnitudes (CLBP_M). The implementation of CLBP_S and CLBP_M are displayed in Figure 5.
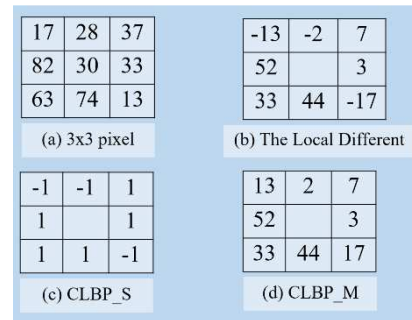
Fig. 5 CLBP Operator

4) Rotated Local Binary Pattern: Rotated LBP (RLBP) [43], sometimes called Dominant Rotated LBP (DRLBP) [44] is a rotation technique on LBP around the center pixel. When the reference is in the circular neighborhood token by dominant direction, then the weights are assigned with reference to the dominant direction. Figure 6 shows the implementation of RLBP.

(a) The 100x100 pixel 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
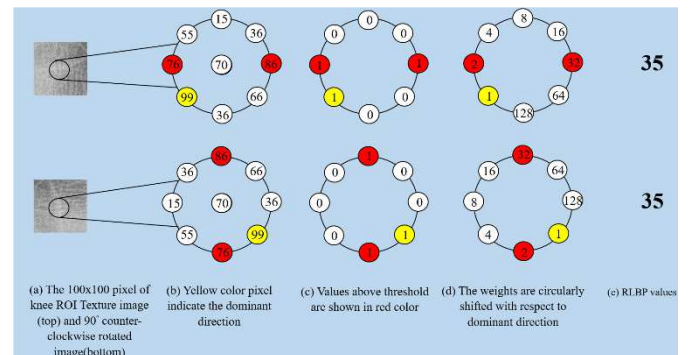
Fig. 6 RLBP Operator

The RLBP is defined by the equation:

$$RLBP_{P,R} = \sum_{P=0}^{P-1} S(g_p - g_c) \rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows 2^{mod(P-D,P)} \qquad (10)$$

where *mod* refers to the operation of modulus, g_p indicates the index of the neighbor pixel, g_c indicates the index of the center pixel, *D* is the dominant direction (D) in a neighborhood that can be written as the equation (11):

$$D = argmax \, | \, g_p - g_c|. \rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows$$
$$\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows, p \in$$
$$(0,1,\ldots,P-1) \qquad (11)$$

5) Rotated Local Binary Pattern : LBP$_{ri}$ is the invariant rotation feature which is based on LBP. With reference to LBP operator, 2^P different output values are the production of LBP. In other words, the 2^P different binary patterns that could be defined by the P pixels in the neighbor set. In the research study presented we have applied LBP$_{ri}$ to each pixel with 8 neighbors, LBP$_{ri}$ with 8 bin is defined by Equation 12:

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

6) Local Binary Pattern Histogram Fourier : Local Binary Pattern Histogram Fourier or LBP-HF descriptor considers as a rotation-invariant technique on image texture that depends on LBPs uniform. The LBP-HF can be defined by the first computation of a non-invariant from LBP histogram over the whole region of images, and then invariant features rotate from the histogram which was constructed. LBP-HF is generally used for static features which used Fast Fourier Transform (FFT) to calculate global features from uniform LBP histogram instead of calculating invariant at each pixel independently. This makes the LBP$_{ri}$ feature set a subset of LBP-HF. Figure 7 shows the implementation of LBP-HF.

Fig. 7 LBP-HF operator for rotation invariant image description

With respect to Figure 7,
If α=45∘, local binary pattern
    00000010 ⇒ 00000100
    00000100 ⇒ 00001000,...,
    11111000 ⇒ 11110001,...,
Similarity if α = K * 45∘, as a consequence, the pattern have to be circularly rotated with k steps.

7) Local Configuration Pattern : Local Configuration Pattern or LCP is an image rotation invariant texture description technique. LCP decomposes the information architecture of an image in two states: (i) microscopic configuration (MiC) and (ii) local structural information. The information consists of image configuration and pixel-wise interaction relationships [45]. While local structure information is directly related to the basic functionality of LBP, MiC is used for exploring microscopic configuration information. The local structure concept implementation is shown in Figure 8.

Fig. 8 LCP Operator

From Figure 8, it pictured that Figures 8(a) and 8(b) are considered to be of the same pattern type as LBP, but with the implementation of LBP with local invariant information.

The patterns illustrated in Figures 8(a) and 8(b) are different, while the patterns are presented in Figures 8(b) and 8(c) are of the same pattern type due to the same value of variance. In contrast, Figures 8(b) and 8(c) are distinct in terms of MiC, which MiC based on textural properties. Microscopic configuration information is defined as the modelling of microscopic configuration as expressed in Equation 13.

$$E(a_0, \ldots, a_{P-1}) = \left| g_c - \sum_{i=0}^{P-1} a_i\, g_i \right| \qquad (13)$$

where g_c and g_i are intensity center pixel values and neighboring pixels; a_i *(i=0,...,P-1)* refers to the weighting parameters associated with g_i ; *E(a$_0$,...,a$_{P-1}$)* refers to the reconstruction error regarding model parameters of a_i

8) Local Ternary Pattern : The Local Ternary Pattern (LTP) is based on improving the basic functionality of LBP, which is the analysis of the central pixel im that tends to be sensitive to noise particularly. With respect to LBP operator presented in Equation 8, *s(x)* contained two values. On the contrary, LTP can be calculated in gray-level in a zone of width ± t. Thus *s(x)* of LBP is replaced by 3-valued function *s$_t$(u,i$_m$,t)* as bellow.

$$S(x) = \begin{cases} 1 & u \le i_m - t \\ 0 & |u - i_m| < t \\ -1 & u \ge i_m + t \end{cases} \qquad (14)$$

9) Haralick Feature : Haralick feature is measured from the Gray-Level Co-occurrence Matrix (GLCM) that is the ordinary way to describe image texture. Haralick feature has been divided into 14 features which are calculated from the statistic of the basic GLCM functionality including.

- *Angular Second Moment (ASM)* is used to find the local uniformity of the grey levels.

$$ASM = \sum_{i=1}^{N}\sum_{j=1}^{N}(P(i,j)^2) \qquad (15)$$

- *Contrast* is a measure grey level variation between the neighbor reference pixel and reference pixel.

$$Contrast = \sum_{i=1}^{N-1} n^2 \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j) \quad, |i-j| = n \quad (16)$$

- *Correlation* is used to show the linear dependency of gray level values in the co-occurrence matrix:

$$Correlation = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(ij)P(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \qquad (17)$$

- *Variance* is the square root value of a grey level variants between the reference pixel and its neighbor measurement

$$\sigma^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu)^2 P(i,j) \qquad (18)$$

- *Inverse Different Moment (IDM)* is the same concept to the inverse *difference* feature, but with lower weights for elements that are further from the diagonal.

$$IDM = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{1+(i-j)^2}P(i,j) \qquad (19)$$

- *Sum Average* is the average of normalized grey tone of an image in the spatial domain.

$$SumAverage = \sum_{i=1}^{2N} i \, P_{x+y}(i) \qquad 21$$

- *Sum Variance* refers to the variance of normalized grey tone of image in the spatial domain

$$SumVariance = \sum_{i=1}^{2N} (i - f_8)^2 P_{x+y}(i) \qquad (21)$$

- *Sum Entropy (f8)* is a measure of randomness within an image

$$SumEntropy = \sum_{i=1}^{2N} \quad P_{x+y}(i) \, log \, P_{x+y}(i) \qquad (22)$$

- Entropy refers to the indication of the complexity within an image

$$Entropy = -\sum_{i=1}^{N} \sum_{j=1}^{N} P(i,j) \, log[\, P(i,j)] \qquad (23)$$

- *Different variance* is an image variation in a normalized co-occurrence matrix

$$DifferenceVariance = \sum_{i=1}^{N} \quad i^2 P_{x-y}(i) \qquad (24)$$

- *Different entropy* is an indication of the amount of randomness in an image

$$DifferenceEntropy = -\sum_{i=1}^{N} \quad P_{x-y}(i) \, log[\, P_{x-y}(i)] \quad (25)$$

- Information *Measure of Correlation 1 (IMC1)* is estimated using two different measures

$$IMC1 = \frac{HXY - HXY1}{max(HX,HY)} \qquad (26)$$

where *HXY* is the value of Entropy; *HX* and *HY* are the entropy of P_x and P_y

$$HXY1 = -\sum_{i=1}^{N} \sum_{j=1}^{N} \quad P(i,j) \, log(\, P_x(i)P_y(j)) \quad (27)$$

- Information *Measure of Correlation 2 (IMC2)*

$$IMC2 = \sqrt{1 - exp[\, -2(HXY2 - HXY)]} \qquad (28)$$

where

$$HXY = -\sum_{i=1}^{N} \sum_{j=1}^{N} \quad P(i,j) \, log(\, P \quad (i,j)), \qquad (29)$$

$$HXY1 = -\sum_{i=1}^{N} \sum_{j=1}^{N} \quad P_x(i)P_y(j) \, log(\, P_x(i)P_y(j) \, (30)$$

- Maximum *Correlation Coefficient (MCC)*

$$MCC = \sqrt{\sum_{k=1}^{N} \frac{P(i,k)P(j,k)}{P_x(i)P_y(j)}} \qquad (31)$$

10) Gabor Filter Feature : Gabor filter is another texture extraction technique which is used to analyze texture for specific local regions of images with the specific frequency and specific direction. Gabor filter is in fact part of the 2D Gabor filter bank that comprises different elements for example frequencies, orientations, and smooth parameters of Gaussian envelope. In addition, Gabor filter bank of pixel (x, y) can be obtained by the Equation 30 below.

$$G(x,y) \equiv e^{-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}} \, e^{j(\omega_{x_0} x + \omega_{y_0} y)} \qquad (32)$$

where ω_{x0} and ω_{y0} are the center frequency of *x* and *y* direction. σ_x and σ_y are the Gaussian function standard deviation along *x* and *y* direction.

E. Feature Selection

In order to get good feature vectors that can be applied to classification process, the feature selection is proposed. Feature selection is one of the major processes in this research work as it involves the selection of useful features in order to reduce the data dimensionality for the classification process. There are five different feature selection methods used in this work: (i) Correlation-based Feature Selection (CFS), (ii) Chi-Square, (iii) Information Gain, (iv) Gain Ratio, and (v) Relief. Each feature selection technique is described in the following sub-sections.

1) Correlation-based Feature Selection: Correlation-based Feature Selection or CFS is a well-known feature selection method that uses a search of a heuristic for evaluating the worth of subsets of features. In other words, CFS is used to calculate subsets for the evaluation of features with the following of the basic hypothesis which are based on the heuristic that "Good feature subsets contain features highly correlated with the classification, yet uncorrelated to each other" [46], [47]. In addition, in CFS, it is applied Symmetric Uncertainty, which is the technique used to reduce the redundancy of a feature. The Symmetric Uncertainty, which applies to two nominal attributes A and B is given by the equation shown below.

$$U(A,B) = 2 \frac{H(A) + H(B) - H(A,B)}{H(A) + H(B)} \qquad (33)$$

where *H* stands for the entropy function. *H(A, B)* stands for the joint entropy of *A* and *B*. The value of symmetric uncertainty can start from 0 till 1. With respect to Equation 32, CFS can be defined as:

$$CFS = 2 \frac{\sum_{j=1}^{N} U(A_j, C)}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} U(A_i, A_j)}} \qquad (34)$$

where *C* refers to the class of feature; *(Ai, Aj)* stand for a pair of attributes in the features set.

2) Chi-Square (χ2): In the literature, one way to measure the dependency between a feature and a class is by using Chi-Square (χ2) feature selection [48]. In the implementation of Chi-Square (χ2), the features are ranked from the most to less useful. ChiSquare (χ2) is defined by the function shown below.

$$\chi^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad (35)$$

where O_{ij} refers to the observed frequency. E_{ij} refers to the expected frequency.

3) Information Gain: Information Gain or IG calculates the information in bits for the prediction of the class if the information available refers to the presence of a feature and correlative with class distribution [49]. IG is applied to select the test attribute at each node. In other words, IG is a feature evaluation method which is based on entropy [50]. For example, the IG of a feature t that give a relation to the collection of aspects A is written in Equation 35.

$$IG(A,t)$$
$$= Entropy(A) - \sum_{v \in Values(t)} \frac{|A_v|}{|A_t|} Entropy(A_v) \qquad (36)$$

where *Values(t)* represents the set of all feature *t* possible values; A_v represents the subset of S with aspects of class v connected to feature *t*; A_t represents the set of all aspects belonging to feature *t*; |·| refers to the cardinality of a set.

4) Gain Ratio: Gain Ratio (GR) is the update or correction of Information Gain (IG). Information Gain is used in a decision tree to select the test attribute at each node [51]. Hence, the implementation of GR to reduce IG bias while choosing an attribute by taking the number and size of branches. GR of an attribute (attr) is defined in the equation below:

$$GR(attr) = \frac{Gain(attr)}{Entropy(attr)} \qquad (37)$$

5) Relief: The final feature selection technique considered in this research refers to relief, which is a weight-based algorithm where the relevant features are the ones that have a better distinction between the classes [52]. Relief works by taking a dataset with n instances of p features which belong to two known classes, each feature of the dataset is scaled to the interval [0 1]. Relief will be repeated m times. It starts with a p-long weight vector (W) of zeros. In other words, relief can compute two feature weights: (i) the near-hit score and (ii) the near-miss score, which is based on nearest instances in the neighborhood. At each repetition of m times, the algorithm selects the feature vector (X) which belongs to a random instance, and the feature vectors of that instance closest to X from each class. In other words, the 'Near-hit' refer to the closest same class of the instance, while the 'Near-miss refers to the closest different-class of instance. After m iterations are finished, the algorithm divides each element of the weight feature vector by m. The creation of relevance vector is accrued. Thus, the features are selected if their relevance is greater than a threshold T.

F. Classification

The classification is used to identify to which class (OA or non-OA) a new knee X-ray image belongs based on a previously developed classification model. The classification algorithms adopted in this research are (i) Decision Tree (C4.5), (ii) Binary Split Tree, (iii) Average One-Dependence Estimators (AODE), (iv) Bayesian Network (BN), (v) Naïve Bayes Classifier, (vi) Support Vector Machine (SVM), (vii) Logistic Regression, (viii) Sequential Minimal Optimization (SMO), and (ix) Backpropagation. Each classification algorithm is described in sub-sections as follow.

1) Decision Tree : Decision tree learning (C4.5) is a well-known learning algorithm used in machine learning [53]. A classification decision tree comprises of the following parts: (i) root node, (ii) internal nodes, and (iii) leaf nodes. The root node is the main root or parent node and has no incoming edges. Internal nodes are used to test on an attribute, and the branch defines the output of the test. The leaf nodes, which represent the classes, are the nodes that are at the bottom of the tree and that have no outgoing edges.

2) Binary Split Tree: The binary split tree is the tree in which each node of the tree comprises only two values (binary 0-1). The split tree is designed for storing statistic datasets with skewed frequency distributions. The split tree is of the type of decision tree while each node of a decision tree can have multiple values. The split value of the split tree is applied for further search in the tree when the key value is not meet to the search value.

3) Average One-Dependence Estimators (AODE): Average One-Dependence Estimators (AODE) is a probabilistic classification learning technique which improves the Naïve Bayesian classifier [54] by addressing the problem of attribute-independence. For instance, in the class y, which has a set of features x1,..., xn, AODE can be used to find the probability of each class y by using the following equation:

$$\hat{P}(y|x_1, \dots, x_n)$$

$$= \frac{\sum_{i:1 \le i \le n \wedge F(x_i) \ge m} \hat{P}(y, x_i) \prod_{j=1}^{N} \hat{P}(x_i|y, x_i)}{\sum_{y' \in Y} \sum_{i:1 \le i \le n \wedge F(x_i) \ge m} \hat{P}(y', x_i) \prod_{j=1}^{N} \hat{P}(x_i|y', x_i)} \qquad (38)$$

where $\widehat{P}$ is represented an estimate of *P*; *F* has represented the frequency; *m* has represented a user specified minimum frequency.

4) Bayesian Network: Bayesian Network (BN) or Probabilistic Networks (PNs) is a graphical probability model used for reasoning and decision making in uncertainty [55]. In other words, the Bayesian network is considered as the directed acyclic graph (DAG), and each node n ∈ N of BN stand for a domain variable or dataset attribute. In addition, the Bayesian network highly depends on the Bayes rule. The Bayes' rule can be written as follows: Assume Ai attribute where i= 1,...,n and take value ai 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:

$$P(C|U) = argmax \, P\,(C)P(U|C) \qquad (39)$$

5) Naïve Bayes Classifier: Naïve Bayes consider one of the most well-known Bayesian techniques; it uses a simple probabilistic classifier with reference to the implementation of Bayes' theorem combined with strong or naive independence assumptions [56]. In the related work of Naïve Bayes learning classifier, it has been assumed that all the attributes in the same class have been considered as independently given a class label. With respect to Bayes rule, the Naïve Bayes classifier has been modified as the equation below:

$$P(C|U) = argmax \, P\,(C) \prod_{i=1}^{N} P(A_i|C) \qquad (40)$$

6) Support Vector Machine: Support Vector Machine or SVM is a popular linear classifier that has been widely used for the classification task. The objective of SVM is to separate instances of two classes in the most optimal way by constructing an N-dimensional hyperplane between two training sample classes in the feature set [57]. SVM classifiers are grouped into two sections: (i) linear and (ii) non-linear. They are explained in the following sub-sub-sections.

- *Linear Classification:* In the context of linear classification, the SVM can be divided into two types of classification: (i) linearly separable case, and (ii) non-linearly separable case. In the linearly separable

case, SVM with the training data x_i, y_i, $y_i \square\text{-}1, +1$, $i = 1,...,n$ can be defined as shown in the equation below.

$$\begin{cases} x_i.\omega + b \geq +1; y_i = +1 \\ x_i.\omega + b \geq -1; y_i = -1 \end{cases} \quad (41)$$

- *Non-Linear Classification:*

In the case of non-linear classification, the SVM equation can be written as:

$$f(x) = \sum_{i=1}^{n_s} \alpha_i y_i P(x_i, x) + b \quad (42)$$

where n_s refers to the number of support vector. α is non-negative Lagrange multipliers; $P(x, y)$ is Polynomial of degree m: $k(x,y)=(x.y+1)^m$

7) Logistic Regression: Logistic regression is a well-known statistical regression model and it is based on ordinary regression [58]. In this research work, the logistic regression has been applied to the dependent variable. The purpose of logistic regression is to discover the best fitting model for evaluating the relationship between a set of independent variables (predictor) and the dichotomous characteristic of interest (Outcome variable). Logistic regression provides the formula to predict a logit transformation of the probability as the Equation 42, where p is the probability of presence of the characteristic of interest.

$$\text{logit(p)} = b0 + b1X1 + b2X2 + b3X3...+ ... bkXk \quad (43)$$

8) Sequential Minimal Optimization: Sequential Minimal Optimization or SMO is the improvement from SVM to find the solution of the quadratic programming (QP) optimization problem, which happens during the SVM training [59]. In order to get the solution from the SVM QP problem, SMO decomposes SVM QP problem into QP sub-problems then solves the smallest possible optimization problem which involves two Lagrange multipliers, at each step. By applying Lagrangian, the QP problem can be transformed into a dual where the objective function Ψ consider as solely dependent on a set of Lagrange multiplier α_i. The equation of Lagrange is presented in Equation 44.

$$min \rightleftarrows \rightleftarrows \psi \rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows (\vec{\alpha})$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j\ y_i y_j x_i x_j \quad (44)$$

9) Backpropagation Neural Network: The backpropagation algorithm is widely known and considered as neural networks model for a gradient calculation which is required in the calculation of the weight to be used in the network [60]. In the Neural Network (NN) has three layers: (i) input layer, (ii) hidden layer and (ii) output layer. Backpropagation is used to train in neural for learning. In other words, in the case of learning in NN, backpropagation is commonly applied by the gradient descent optimization algorithm in order to adjust the weight of neurons by measuring the gradient of the loss function. This method is sometimes known as backward propagation of errors because the error is measured at the output layer and distributed back through the network layers.

In other words, the backpropagation neural network known as the Multilayer Perceptions grouped neurons into layers. For the input layer and output, layer are represented by the first layer and the last layer respectively. The hidden layers are the remaining layers of the network. The back propagation neural network is illustrated in Figure 9.
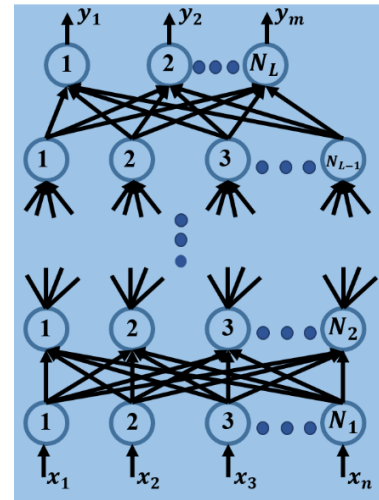

Fig. 9 Backpropagation Neural Network

III. RESULT AND DISCUSSION

In this section, the evaluation of the OA screening considered as the study proposed approach is illustrated. While 7,272 experiments were handled with respect to the proposed approach, only the most significant results obtained are presented. The study evaluation handled by considering a case study with 131 digital medical X-ray imageries taken in Postero-Anterior (PA) position. The dataset comprised: (i) X control (normal) and (ii) Y OA images. TenFold Cross-Validation (TCV) was implemented in the study. The evaluation measures used were: (i) Area Under the Receiver Operating Characteristic Curve (AUC), (ii) Accuracy (AC), (iii) Sensitivity (SN), (iv) Specificity (SP), (v) Precision (PR), and (vi) F-Measure (FM). The overall aim of the study evaluation was to prove some evidence that the OA be able to detected by applying the proposed framework. Finally, there are four groups of experiments were handled with the objectives of comparing and selecting the best results for the following criteria:

- Region of Interest (RoI) is the set of the experiments conducted in order to investigate the most appropriate RoI.
- Texture descriptor is the set of the experiment performed to find the most appropriate texture descriptor method.
- Feature selection technique is the collection of the experiments conducted to identify the most appropriate feature selection technique.
- Classification algorithm is the collection of the experiments performed in order to get the most appropriate classification learning algorithm.

The comparison and results for each area of interest are presented in the following sub-sections.

A. Region of Interest (RoI)

This sub-section discusses the evaluation conducted to compare the best result of applying different four RoIs: (i) Medial Femur (MF), (ii) Lateral Femur (LF), (iii) Medial

Tibia (MT) and (iv) Lateral Tibia (LT). In the research experiment, LBP descriptor was applied with CFS feature selection (LBP and CFS feature selection were used because the reports in Sub-section B and Sub-section C, had revealed that this was appropriate texture descriptor and feature selection, respectively) and a Bayesian Network classifier method as these had been found to work well in the context of OA detection (see Sub-section D). The best performance of each RoI is shown in Table 1 below (best result indicated in bold font with respect to AUC values):

TABLE I
REGION OF INTEREST RESULTS

RoI	AUC	AC	SN	SP	PR	FM
MF	0.884	0.794	0.794	0.792	0.794	0.794
LF	**0.912**	**0.832**	**0.832**	**0.832**	**0.832**	**0.832**
MT	0.895	0.802	0.802	0.802	0.802	0.802
LT	0.883	0.809	0.809	0.809	0.809	0.809

From Table 1, it can be concluded that Literal Femur (LF) is the most appropriate one amount of four RoIs of texture analysis for OA detection which came with the best value of AUC of 0.912, while the second-highest appropriate RoI went to Medial Femur (MF) with the AUC value of 0.895. It should be suggested that Lateral Femur is first selecting an area for texture analysis of OA detection.

B. Texture Descriptor

The evaluation conducted to identify the best algorithm for feature extraction (Texture Descriptor) of the sub-image are presented in this sub-section. Ten feature extraction algorithms were considered: (i) histogram feature, (ii) Local Binary Pattern, (iii) Completed LBP, (iv) Rotated Local Binary Pattern, (v) Local Binary Pattern Rotation Invariant, (vi) Local Binary Pattern Histogram Fourier, (vii) Local Ternary Pattern, (vii) Local Configuration Pattern, (ix) Haralick feature, and (x) Gabor filter feature descriptor. For the experiment used to compare these 10 algorithms the LF RoI was used as this had been found to produce the best result (presented in the previous sub-section). Again CFS feature selection was adopted together with Bayesian Network classifier for the same reason as before. The best performance of each texture descriptor is illustrated in Table 2 below.

TABLE II
TEXTURE DESCRIPTOR RESULTS

Texture Descriptor	AUC	AC	SN	SP	PR	FM
Histogram	0.757	0.695	0.695	0.690	0.695	0.693
LBP	**0.912**	**0.832**	**0.832**	**0.832**	**0.832**	**0.832**
CLBP	0.882	0.763	0.763	0.762	0.763	0.763
RLBP	0.895	0.809	0.809	0.810	0.810	0.809
LBP$_{ri}$	0.812	0.771	0.771	0.771	0.771	0.771
LBP-HF	0.773	0.710	0.710	0.717	0.710	0.709
LTP	0.816	0.756	0.756	0.761	0.763	0.755
LCP	0.783	0.725	0.725	0.724	0.725	0.725
Haralick	0.695	0.664	0.664	0.670	0.672	0.662
Gabor	0.883	0.786	0.786	0.786	0.786	0.786

From Table 2, it can be observed that Local Binary Pattern (LBP) is considered as the best texture descriptor for OA detection work combine with LF RoI can produce the best AUC value of 0.91. For the second-best texture descriptor performance went to Rotated Local binary Pattern (RLPB) which one of the extension techniques from LBP, the best result of RLBP produced the second-highest of ACU value of 0.895 in case of texture analysis of OA detection amount of four RoIs. Based on Table II, it can be suggested that LBP is the first choice for using in OA detection and the second choice went to RBLP. On the other hands, Haralick feature produced the lowest result of AUC compare to other texture descriptors. It can be suggested that Haralick is the last choice in this case.

C. Feature Selection Technique

In this sub-section, the evaluation conducted to determine the best mechanism for dimension reduction of the feature vector of each sub-images is described. Five algorithms of feature selection were applied: (i) Correlation-based Feature Selection (CFS), (ii) Chi-Squared, (iii) information gain, (iv) Gain Ration, and (v) Relief feature selection. In the experiments used to observe these five mechanisms the LF RoI and the LBP descriptor were used as this had been found to produce the best result was presented in the previous sub-sections (Sub-section A and B). Again the implementation of Bayesian Network classifier for the same reason as before. The best performance of each feature selection mechanism is presented in Table 3 below:

TABLE III
FEATURE SELECTION TECHNIQUE BEST RESULTS

Feature Selection	AUC	AC	SN	SP	PR	FM
CFS	**0.912**	**0.832**	**0.832**	**0.832**	**0.832**	**0.832**
χ2	0.699	0.687	0.687	0.687	0.687	0.687
GR	0.709	0.687	0.687	0.687	0.687	0.687
IG	0.699	0.687	0.687	0.687	0.687	0.687
Relief	0.699	0.679	0.679	0.674	0.681	0.677

In Table 3, Correlation Based Feature selection (CFS) is the best Texture selectors which applied with LF and LBP to produce the highest value of AUC at 0.912. Gain ration is the second best of texture selector which can produce the value of AUC at 0.709. On the contrary, Chi-squared, Information gain and relief produced the same value of AUC with the value of 0.699. In term of AUC value, it can be suggested that CFS is the first texture selector for knee OA detection applied with LF RoI and LBP texture analysis technique.

D. Learning Algorithm

The reports on the evaluation conducted to analyze the best mechanism for generation classifier of learning methods is presented in this sub-section. Nine algorithms of learning methods were considered for the experiment: (i) Decision Tree, (ii) Binary Split Tree, (iii) Average One-Dependence Estimators, (iv) Bayesian Network, (v) Naïve Bayes, (vi) Support Vector Machine, (vii) Logistic regression, (vii) Sequential Minimal optimization, and (ix) Back Propagation Neural Network. For the experiments used to compare these

nine mechanisms the LF RoI and the LBP descriptor were used as this and applied with CFS feature selection had been found to produce the best result was presented in the previous sub-sections (Sub-section A, B and C). The best performance of each learning algorithm is illustrated in Table 4.

TABLE IV
LEARNING ALGORITHM BEST RESULTS

Learning Algorithm	AUC	AC	SN	SP	PR	FM
Decision Tree	0.757	0.779	0.779	0.780	0.780	0.779
Binary Split Tree	0.766	0.740	0.740	0.736	0.742	0.739
AODE	0.896	0.809	0.809	0.804	0.809	0.809
Bayesian Network	**0.912**	**0.832**	**0.832**	**0.832**	**0.832**	**0.832**
Naïve Bayes	0.903	0.817	0.817	0.816	0.817	0.817
SVM	0.715	0.718	0.718	0.711	0.720	0.715
Logistic Regresion	0.904	0.840	0.840	0.844	0.847	0.839
SMO	0.771	0.771	0.771	0.771	0.771	0.771
Neural Network	0.851	0.771	0.771	0.770	0.771	0.771

From Table 4, it shows that Bayesian Network is the best learning method that can produce the highest value of AUC with the value of 0.912, while the second-best learning method went to Logistic regression with the AUC value of 0.904. In contrast, support vector machine is the lowest learning method for selection in case of OA detection due to the production of AUC value of 0.715 that considered as the lowest AUC value amount of learning methods applied in the study. In short, it should be suggested that the applying of Bayesian network to LF RI, LBP texture descriptor, and CFS feature selection approach produced the highest AUC value of 0.912.

IV CONCLUSION

The framework of early detection of OA by applying image processing and classification techniques to human knee X-ray imagery was presented. The detection was applied with the difference of four sub-images: (i) Lateral Femur (LF), (ii) Lateral Tibia (LT), (iii) Medial Femur (MF), and (iv) Medial Tibia (MT). In addition, in order to carry out a more comprehensive research study, ten types of texture descriptors were applied to the sub-imagery of X-ray images, for example (i) Histogram feature, (ii) Local Binary Pattern (LBP), (iii) Completed LBP (CLBP), (iv) Rotated LBP (RLBP), (v) LBP Histogram Fourier (LBPHF), (vi) LBP Rotation Invariant (LBPri), (vii) Gabor, (viii) Haralick, (ix) Local Configuration Pattern (LCP), and (x) Local Ternary Pattern (LTP). The implementation of feature selection was presented to reduce the dimension and space of the features from each feature descriptor. For the feature selection technique, the research study applied five well-known techniques: (i) Correlation-based Feature Selection (CFS), (ii) Chi-square, (iii) Gain Ratio, (iv) Information Gain, and (v) Relief. The last part of the study focused on the implementation of nine machine learning classification algorithms: (i) Decision Tree (C4.5), (ii) Decision with binary tree, (iii) Average One-Dependence Estimators

(AODE), (iv) Bayesian Network, (v) Naïve Bayesian Classifier, (vi) Support Vector Machine (SVM), (vii) Logistic Regression, (viii) Sequential Minimal Optimization (SMO), and (ix) the Backpropagation algorithm. The classification results of the research were evaluated by six different evaluation measures: (i) Area Under the Receiver Operating Characteristic Curve (AUC), (ii) Accuracy (AC), (iii) Sensitivity (SN), (iv) Specificity (SP), (v) Precision (PR), and (vi) F-Measure (FM).

The best classification result obtained had an AUC value at 0.912. With the best result of the research study presented, four main interesting aspects can be listed as follow :

- Regarding the amount of four sub-imagery of knee image, the Lateral Femur (LF) produce a better performance of classification in terms of the implementation of LBP with Bayesian Network.
- In the implementation of the ten texture descriptors, only LBP produced the AUC value at 0.912, which was the highest value recorded in the study.
- The best performance of feature selection for classification was recorded by CFS with five well-known feature selection techniques. – The most efficient learning algorithm for classification was performed by Bayesian Network.
- The highest AUC value of 0.912 was recorded by the implementation of 4 parameters: (i) Lateral Femur sub-image, (ii) LBP descriptor, (iii) CFS feature selector, and (iv) Bayesian Network algorithm.

Future work considered includes using a larger dataset and implementing Convolutional Neural Network (CNN), which is a small branch of Deep Learning that can remove the feature selection process for learning OA/Normal classification imagery.

REFERENCES

[1] J. C. Mora, R. Przkora, and Y. Cruz-Almeida, "Knee osteoarthritis: Pathophysiology and current treatment modalities," *J. Pain Res.*, vol. 11, pp. 2189–2196, 2018.

[2] Y. Zhang and J. Jordan, "Epidemyology of Osteoarthritis," *Clin. Geriatr. Med.*, vol. 26, no. 3, pp. 355–369, 2010.

[3] S. S., P. U., and R. R., "Detection of Osteoarthritis using Knee X-Ray Image Analyses: A Machine Vision based Approach," *Int. J. Comput. Appl.*, vol. 145, no. 1, pp. 20–26, 2016.

[4] T. Janvier, R. Jennane, H. Toumi, and E. Lespessailles, "Subchondral tibial bone texture predicts the incidence of radiographic knee osteoarthritis: data from the Osteoarthritis Initiative," *Osteoarthr. Cartil.*, vol. 25, no. 12, pp. 2047–2054, 2017.

[5] M. Kotti, L. D. Duffell, A. A. Faisal, and A. H. McGregor, "Detecting knee osteoarthritis and its discriminating parameters using random forests," *Med. Eng. Phys.*, vol. 43, pp. 19–29, 2017.

[6] M. Camacho-Encina *et al.*, "Discovery of an autoantibody signature for the early diagnosis of knee osteoarthritis: data from the Osteoarthritis Initiative," *Ann. Rheum. Dis.*, p. annrheumdis-2019-215325, 2019.

[7] N. Bayramoglu, A. Tiulpin, J. Hirvasniemi, M. T. Nieminen, and S. Saarakkala, "Adaptive Segmentation of Knee Radiographs for Selecting the Optimal ROI in Texture Analysis," pp. 18–21, 2019.

[8] A. Suponenkovs, Z. Markovics, and A. Platkajis, "Computer Analysis of Knee by Magnetic Resonance Imaging Data," *Procedia Comput. Sci.*, vol. 104, no. December 2016, pp. 354–361, 2016.

[9] J. Ren, Y. Yuan, Y. Shi, and X. Tao, "Tumor heterogeneity in oral and oropharyngeal squamous cell carcinoma assessed by texture analysis of CT and conventional MRI: a potential marker of overall survival," *Acta radiol.*, 2019.

[10] C. Jin, Y. Yang, Z.-J. Xue, K.-M. Liu, and J. Liu, "Automated Analysis Method for Screening Knee Osteoarthritis using Medical Infrared Thermography," *J. Med. Biol. Eng.*, vol. 33, no. 5, p. 471,

2013.

[11] M. George and R. Zwiggelaar, "Comparative study on local binary patterns for mammographic density and risk scoring †," *J. Imaging*, vol. 5, no. 2, 2019.

[12] Z. Guo, L. Zhang, and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, 2010.

[13] V. T. Hoang, A. Porebski, N. Vandenbroucke, and D. Hamad, "LBP parameter tuning for texture analysis of lace images," *IPAS 2016 - 2nd Int. Image Process. Appl. Syst. Conf.*, no. November, 2017.

[14] Z. Wang, R. Huang, W. Yang, and C. Sun, "An enhanced Local Ternary Patterns method for face recognition," *Proc. 33rd Chinese Control Conf. CCC 2014*, vol. 0, no. 1, pp. 4636–4640, 2014.

[15] X. Tan and B. Triggs, "Recognition Under Difficult Lighting Conditions," *IEEE Trans. image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[16] P. P. Kawathekar, "Osteoarthritis of knee joint," pp. 1–4, 2015.

[17] D. Cherezov et al., "Revealing Tumor Habitats from Texture Heterogeneity Analysis for Classification of Lung Cancer Malignancy and Aggressiveness," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.

[18] S. Y. Jeong et al., " Prediction of Chemotherapy Response of Osteosarcoma Using Baseline 18 F-FDG Textural Features Machine Learning Approaches with PCA ," *Contrast Media Mol. Imaging*, vol. 2019, pp. 1–7, 2019.

[19] M. A. Naveed et al., "Grading of oligodendroglial tumors of the brain with apparent diffusion coefficient, magnetic resonance spectroscopy, and dynamic susceptibility contrast imaging," *Neuroradiol. J.*, vol. 31, no. 4, pp. 379–385, 2018.

[20] S. Chinnam, V. Sistla, and V. Kolli, "SVM-PUK Kernel Based MRI-brain Tumor Identification Using Texture and Gabor Wavelets," *Trait. du Signal*, vol. 36, no. 2, pp. 185–191, 2019.

[21] J. H. Park, Y. J. Bae, B. S. Choi, Y. H. Jung, and W. Jeong, "Texture Analysis of Multi-Shot Echo-planar Diffusion-Weighted Imaging in Head and Neck Squamous Cell Carcinoma : The Diagnostic Value for Nodal Metastasis," pp. 1–14, 2019.

[22] H. J. Meyer, G. Hamerla, A. K. Höhn, and A. Surov, "CT texture analysis-correlations with histopathology parameters in head and neck squamous cell carcinomas," *Front. Oncol.*, vol. 9, no. MAY, pp. 1–8, 2019.

[23] K. J. Lafata, Z. Zhou, J. G. Liu, J. Hong, C. R. Kelsey, and F. F. Yin, "An Exploratory Radiomics Approach to Quantifying Pulmonary Function in CT Images," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.

[24] F. W. Feldhaus, D. C. Theilig, R.-H. Hubner, J.-M. Kuhnigk, K. Neumann, and F. Doellinger, "<p>Quantitative CT analysis in patients with pulmonary emphysema: is lung function influenced by concomitant unspecific pulmonary fibrosis?</p>," *Int. J. Chron. Obstruct. Pulmon. Dis.*, vol. Volume 14, pp. 1583–1593, 2019.

[25] R. R. Wildeboer et al., "Automated multiparametric localisation of prostate cancer based on B-mode , shear-wave elastography , and contrast-enhanced ultrasound radiomics," 2019.

[26] M. Zhang et al., "Diagnostic performance of multiparametric transrectal ultrasound in localised prostate cancer: A comparative study with magnetic resonance imaging," *J. Ultrasound Med.*, vol. 38, no. 7, pp. 1823–1830, 2019.

[27] R. Cuocolo et al., "Machine learning applications in prostate cancer magnetic resonance imaging," *Eur. Radiol. Exp.*, vol. 3, no. 1, 2019.

[28] B. Badic et al., "Radiogenomics-based cancer prognosis in colorectal cancer," *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, 2019.

[29] V. Nardone et al., "Magnetic-resonance-imaging texture analysis predicts early progression in rectal cancer patients undergoing neoadjuvant chemoradiation," *Gastroenterol. Res. Pract.*, vol. 2019, pp. 1–9, 2019.

[30] M. del C. V. Hernández et al., "Application of texture analysis to study small vessel disease and blood-brain barrier integrity," *Front. Neurol.*, vol. 8, no. JUL, 2017.

[31] N. Jahani et al., "Prediction of Treatment Response to Neoadjuvant Chemotherapy for Breast Cancer via Early Changes in Tumor Heterogeneity Captured by DCE-MRI Registration," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[32] R. D. Chitalia and D. Kontos, "Role of texture analysis in breast MRI as a cancer biomarker: A review," *J. Magn. Reson. Imaging*, vol. 49, no. 4, pp. 927–938, 2019.

[33] C.-C. Chang, C.-J. Chen, W.-L. Hsu, S.-M. Chang, Y.-F. Huang, and Y.-C. Tyan, "Prognostic Significance of Metabolic Parameters and Textural Features on 18F-FDG PET/CT in Invasive Ductal Carcinoma of Breast," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, 2019.

[34] L. Sannachi et al., "Breast Cancer Treatment Response Monitoring Using Quantitative Ultrasound and Texture Analysis: Comparative Analysis of Analytical Models," *Transl. Oncol.*, vol. 12, no. 10, pp. 1271–1281, 2019.

[35] M. A. M. Shukran, N. S. M. Ahmad, S. Ramli, and F. Rahmat, "Melanoma cancer diagnosis device using image processing techniques," *Int. J. Recent Technol. Eng.*, vol. 7, no. 5, pp. 490–494, 2019.

[36] N. K. El Abbadi and Z. Faisal, "Detection and analysis of skin cancer from skin lesions," *Int. J. Appl. Eng. Res.*, vol. 12, no. 19, pp. 9046–9052, 2017.

[37] E. Mohammed and M. E., "Classification of Dermoscopy Images for Early Detection of Skin Cancer – A Review," *Int. J. Comput. Appl.*, vol. 178, no. 17, pp. 37–43, 2019.

[38] R. Sahoo and C. Sekhar, "Detection of Diabetic Retinopathy from Retinal Fundus Image using Wavelet based Image Segmentation," *Int. J. Comput. Appl.*, vol. 182, no. 47, pp. 46–50, 2019.

[39] A. Imran, J. Li, Y. Pei, J.-J. Yang, and Q. Wang, "Comparative Analysis of Vessel Segmentation Techniques in Retinal Images," *IEEE Access*, vol. 7, pp. 114862–114887, 2019.

[40] L. Yu et al., "Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis," *BMC Cancer*, vol. 19, no. 1, pp. 1–12, 2019.

[41] B. Owen, D. Gandara, K. Kelly, E. Moore, D. Shelton, and F. Knollmann, "CT volumetry and basic texture analysis as surrogate markers in advanced non-small cell lung cancer," *Clin. Lung Cancer*, 2019.

[42] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.

[43] R. Mehta and K. Egiazarian, "Rotated Local Binary Pattern (RLBP): Rotation invariant texture descriptor," *ICPRAM 2013 - Proc. 2nd Int. Conf. Pattern Recognit. Appl. Methods*, pp. 497–502, 2013.

[44] R. Mehta and K. Egiazarian, "Dominant Rotated Local Binary Patterns (DRLBP) for texture classification," *Pattern Recognit. Lett.*, vol. 71, pp. 16–22, 2016.

[45] Y. Guo, G. Zhao, and M. Pietikäinen, "Texture classification using a linear configuration model based descriptor," *BMVC 2011 - Proc. Br. Mach. Vis. Conf. 2011*, pp. 1–10, 2011.

[46] M. Hall, "Correlation-based Feature Selection for Machine Learning," *Methodology*, vol. 21i195-i20, no. April, pp. 1–5, 1999.

[47] M. Lewandowski, D. Makris, and J. C. Nebel, "Automatic configuration of spectral dimensionality reduction methods," *Pattern Recognit. Lett.*, vol. 31, no. 12, pp. 1720–1727, 2010.

[48] K. Pearson, "and the Chi-squared Test," vol. 51, no. 1, pp. 59–72, 2014.

[49] D. Roobaert, G. Karakoulas, and N. V Chawla, "Information Gain , Correlation and Support Vector Machines," *Featur. Extr. Found. Appl.*, vol. 470, no. 2006, pp. 463–470, 2006.

[50] S. Lei, "A Feature Selection Method Based on Information Gain and Genetic Algorithm," *2012 Int. Conf. Comput. Sci. Electron. Eng.*, pp. 355–358, 2012.

[51] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.

[52] K. Kira and L. Rendell, "A practical approach to feature selection," *Proceedings of the Ninth International Conference on Machine Learning*. pp. 249–256, 1994.

[53] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[54] G. I. Webb, J. R. Boughton, and Z. Wang, "Not so naive Bayes: Aggregating one-dependence estimators," *Mach. Learn.*, vol. 58, no. 1, pp. 5–24, 2005.

[55] N. Friedman, K. Murphy, and S. Russell, "Learning the structure of dynamic probabilistic networks," *Proc. Fourteenth* , 1998.

[56] D. Lowd and P. Domingos, "Naive Bayes models for probability estimation," *ICML 2005 - Proc. 22nd Int. Conf. Mach. Learn.*, no. May, pp. 529–536, 2005.

[57] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[58] K. A. Lorenz, "Modeling Binary Correlated Responses using," pp. 17–24, 2015.

[59] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," *Adv. kernel methods*, pp. 185–208, 1998.

[60] H. J. Kelley, "Gradient Theory of Optimal Flight Paths," *ARS J.*, vol. 30, no. 10, pp. 947–954, 1960.