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

**Abstract**

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]. 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 use 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(J 48), ii) Decision with binary true, iii) Average One-Dependence Estimators (AODE), iv) Bayes Network, v) Naïve Bay, vi) Support Vector Machine (libsvm), vii) Logistic, viii) Sequential Minimal Optimization, and ix) Multilayer.

Paper Organization here

1. **Related work**
2. **Proposed Framework**

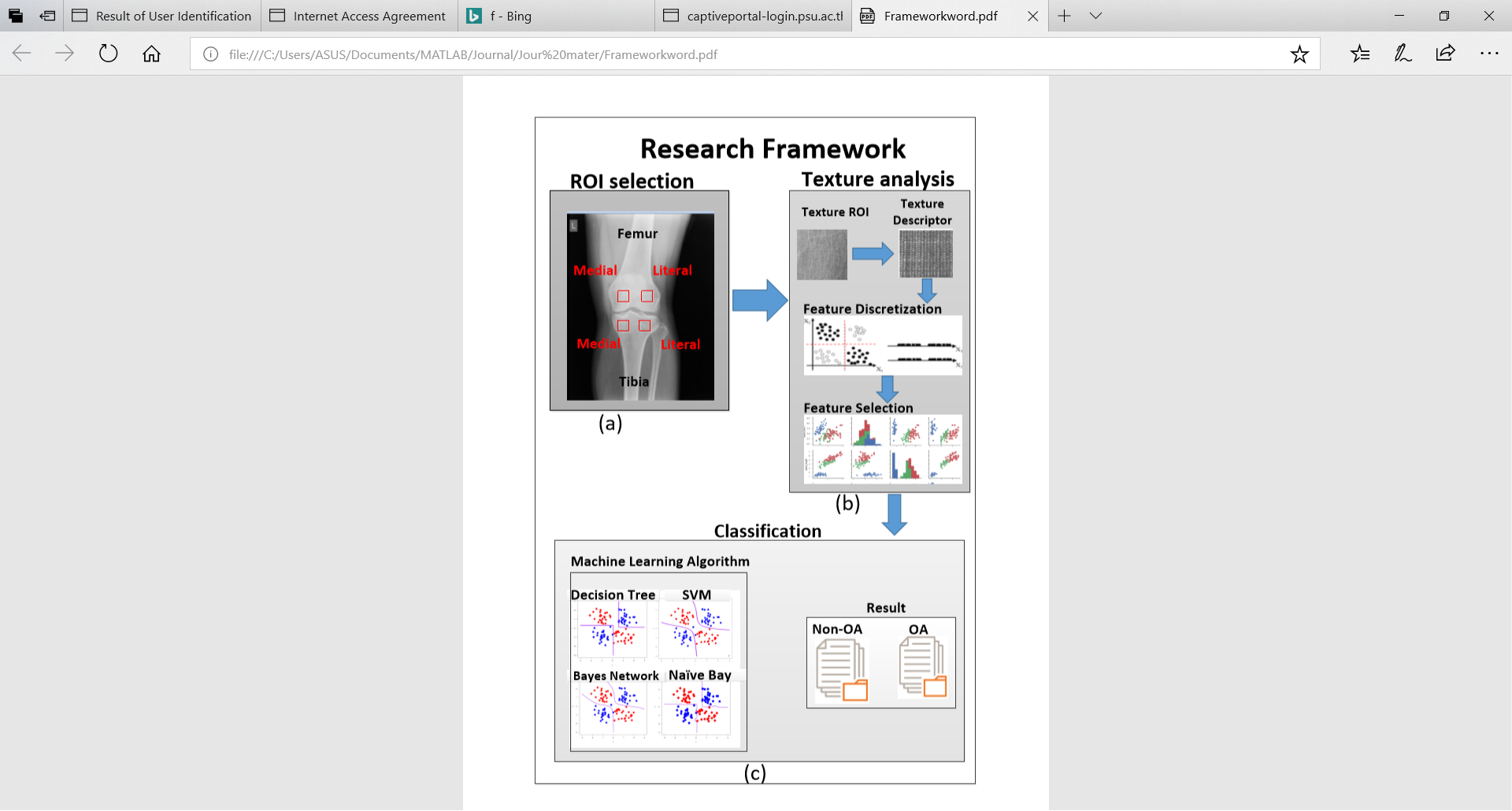


Fig.1.The processes of proposed framework**.**

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.

1. **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:

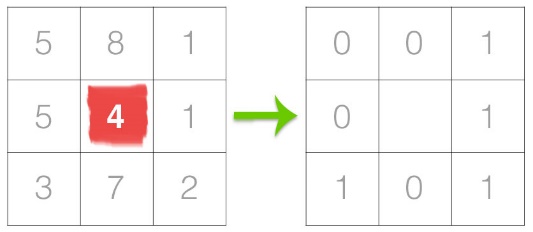
1. **The histogram features**

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

* Mean
* Varian
* Skewness
* Kurtosis
* Energy
* Entropy

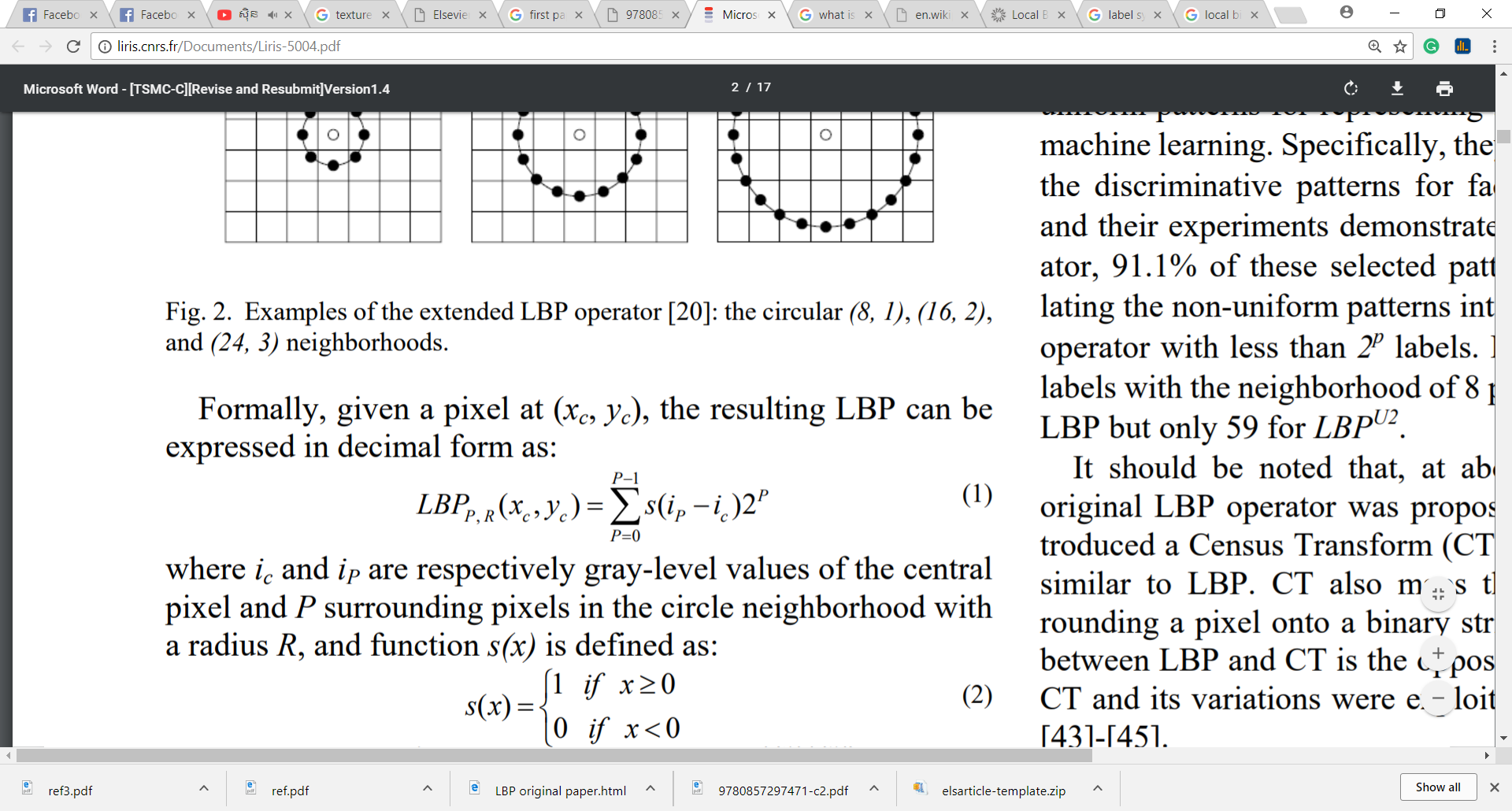
1. **Local Binary Pattern (LBP)**

Local Binary Pattern (LBP) [3] is used to label the pixel which applied thresholding the neighborhood of each pixel with the output as the binary number. The basic of LBP operator is shown in Fig. below:



**Fig. LBP operator**

In addition, LBP at pixe *(xc , yc )* can be calculated by the equation below:



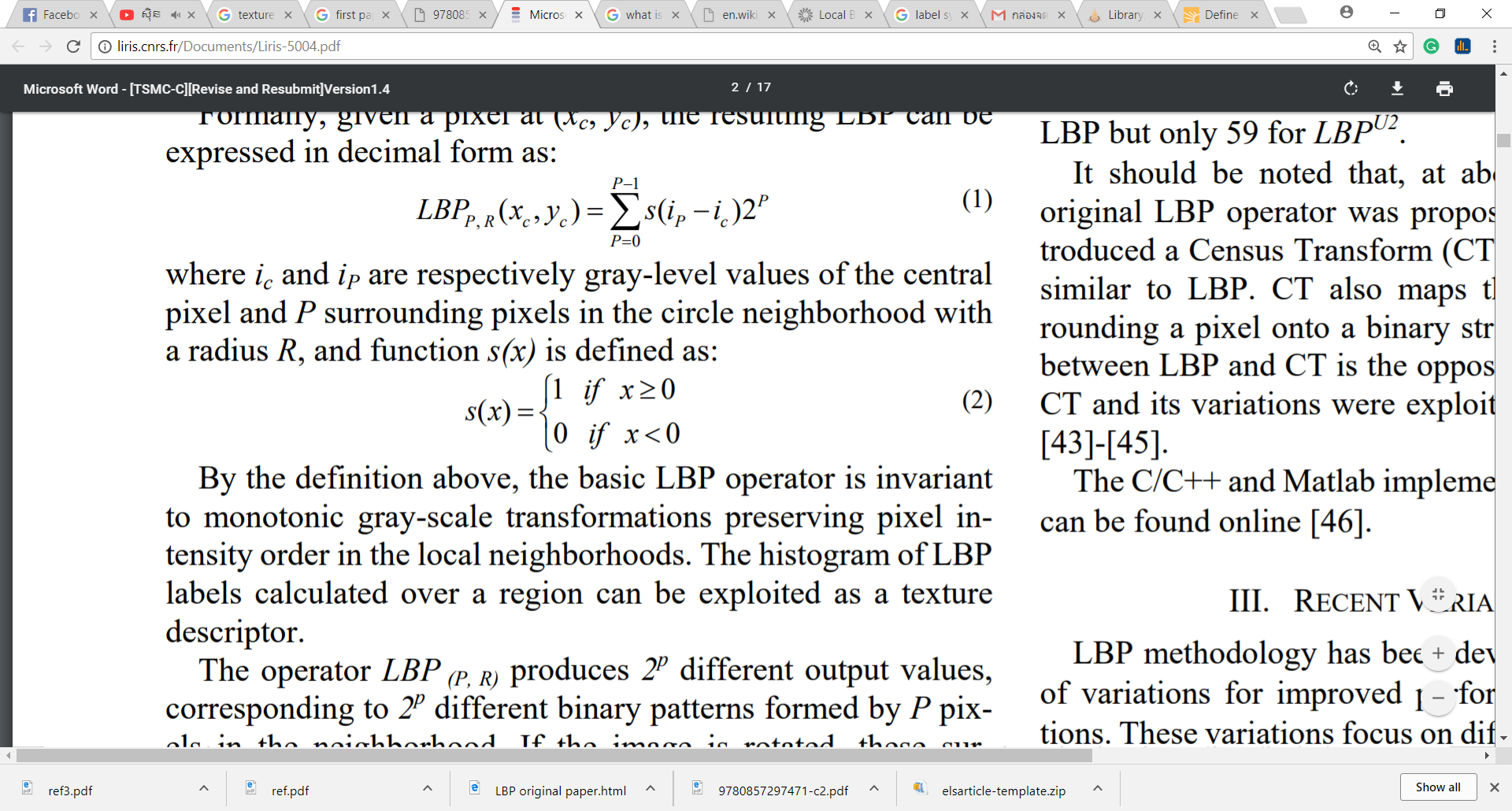
**Where:**

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:



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

1. **Completed LBP (CLBP)**

Completed LBP or CLBP, a local region is defined by center pixel and a local difference sign-magnitude transform (LDSMT). In the research study is focused on LDSMT, LDSCM 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. below:

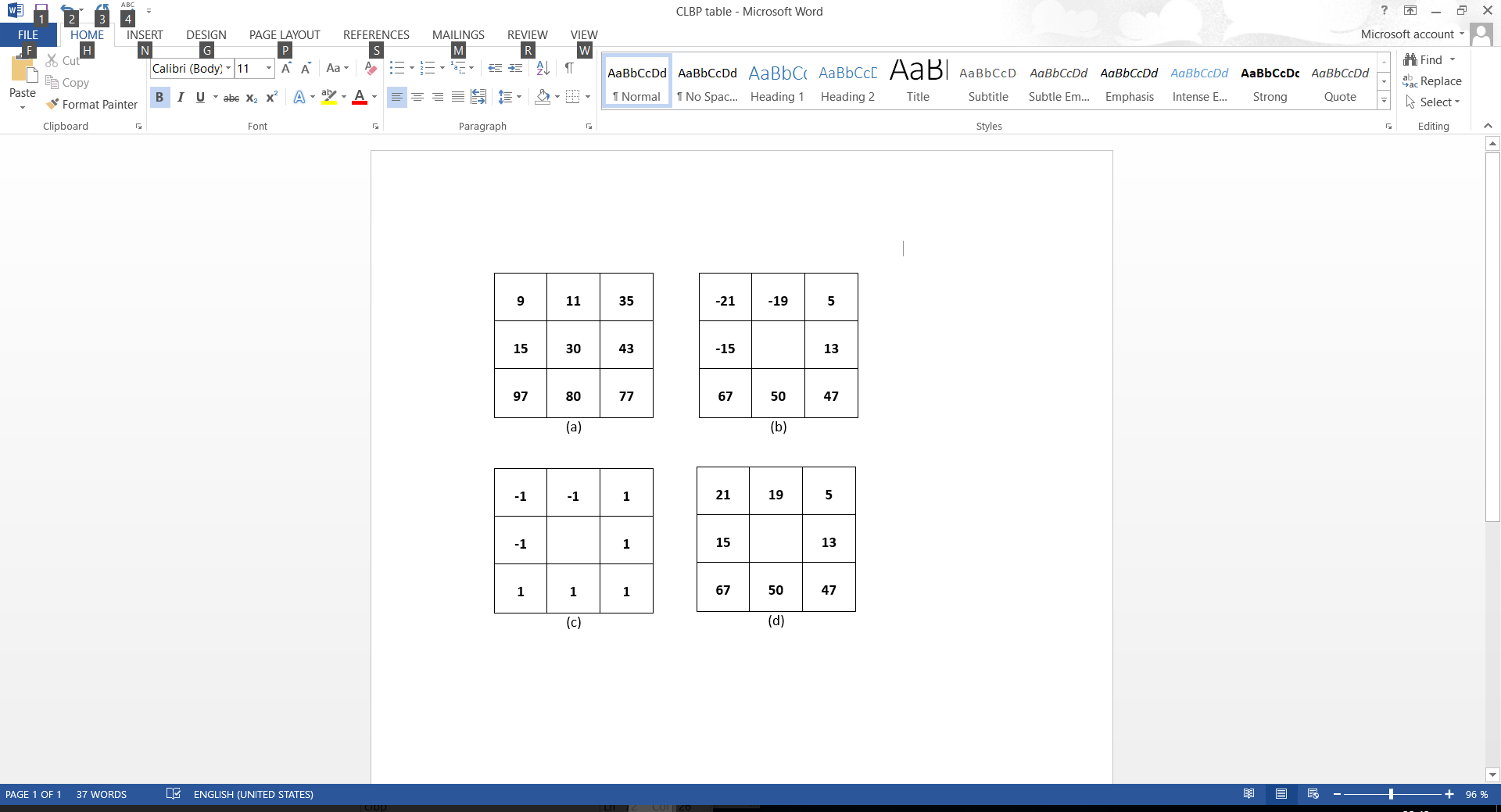


Fig. 2 (a) 3x3 pixel, (b) the local different, (c) CLBP\_S and (d) CLBP\_M

1. **Rotated LBP (RLBP)**
2. **LBP Histogram Fourier (LBP\_HF)**
3. **LBP Rotation Invariant (LBP\_ri)**
4. **Gabor**
5. **Haralick**
6. **Local Configuration Pattern (LCP)**
7. **Local Ternary Pattern (LTP)**
8. **Feature Selection and Classification**

Five difference feature selection techniques are applied in the research work:

1. Correlation-based Feature Selection (CFS), ii)
2. Chi-square
3. Gain Ration
4. Information Gain
5. Relief

**…………………..**

1. Decision Tree(J 48)
2. Decision with binary true
3. Average One-Dependence Estimators (AODE)
4. Bayes Network
5. Naïve Bay
6. Support Vector Machine (libsvm)
7. Logistic
8. Sequential Minimal Optimization (SMO)
9. Multilayer
10. **Data Collection**
11. **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.

1. **Data classify 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. below:

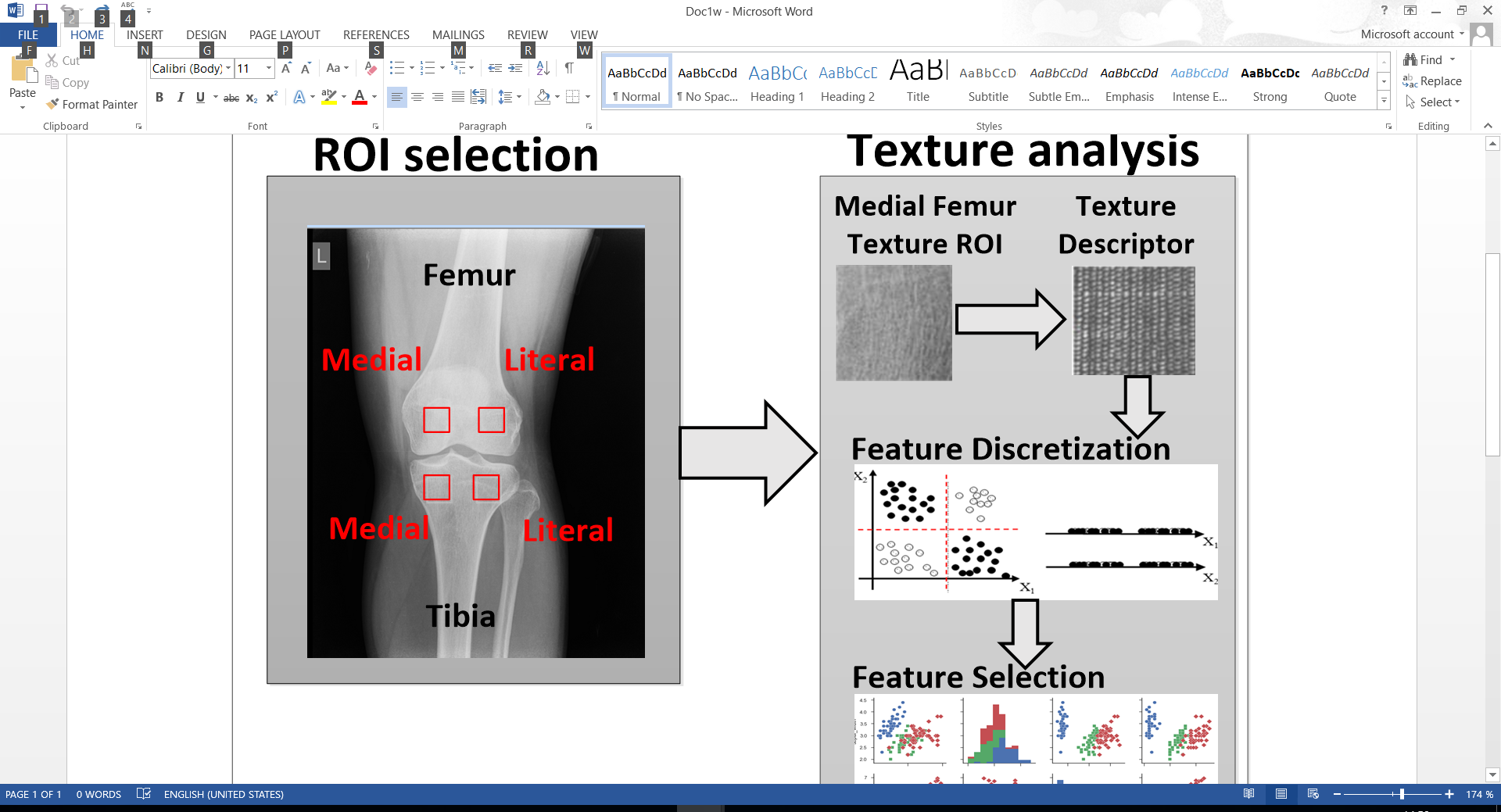


Fig. The ROI of Texture Analysis

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

1. **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: 4 numbers or cases

Feature Descriptor: 10 descriptors

Learning algorithm: 9 algorithms

For feature selection has 5 algorithms for implementation include: CFS, Chi-square, Gain ration, Information Gain, and Relief. On the other hands, Chi-square, Gain ration, Information Gain, and Relief are applied with ranker method, then the best k=10,20,30,40,50, and 60. Hence, all the number of feature selection: CFS (1) + Chi-square (6) + Gain ration (6) + Information Gain (6) + Relief (6) =25 cases.

Finally, all the number of research experiments = 4 x 10 x 9 x 25 = 9000 experiments. The objective of the research experiment is discussed in the next paragraph.

The evolution objective of the research is conducted with four difference objectives include:

1. The Region of Interest (ROI) result comparison: A four different places of ROI for instance: Literal Femur (LF), Medial Femur (MF), Literal Tibia (LT), and Literal Femur (LF) are selected to use for texture analysis to define the OA or non-OA case. This topic is presented detail in subsection 1). The ROI result.
2. The Best feature descriptor result comparison: There are ten of texture descriptor is implemented in the research 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). The result of each feature descriptor is chosen the best result in the research experiment result. The further detail is pointed out in sub-section 2) The Best feature descriptor result..
3. The best feature selection result comparison: In the feature selection process, there are five feature selection techniques is 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 3)The best feature selection.
4. The best learning algorithm result comparison: 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) Naïve Bay, vi) Support Vector Machine (SVM), vii) Logistic, viii) Sequential Minimal Optimization, and ix) backpropagation algorithm.

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

1. **The ROI result**

With the four different ROI: LF, MF, LT, and MT with the best result of each ROI as the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 |

Table1. 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 Naïve Bay, while the best result of LT is given by Gabor applied with Bayes Network for AUC values and other highest values is 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.

1. **The Best feature descriptor result**

In the Literal Femur produce the best result with accuracy, in this point the best feature descriptors are being used to find the best one

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm  Texture  Descriptor | AUC | AC | SN | SP | PR | FM |
| 1st Level GLCM | 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 | 0.77 | 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 |

Table2. Texture Descriptor Result Comparison

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 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 produce by MT applied with Bayes Network. In addition, the best result of Gabor is given by LT applied with Bayes and MT is applied with 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 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 Naïve 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 Naïve bay, and others are given by LT applied with Bayes Network and Naïve Bay.

1. **The Best feature selection**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm  Feature  Selection | 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 Naïve 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 is pointed out by LT applied with LTP and SMO. In the information gain best result is presented by LF applied with GLCM and Naïve 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 Naïve Bay. The best result of learning algorithm is presented in next subsection.

1. **The Best learning algorithm**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Algorithm  Learning  Algorithm | AUC | AC | SN | SP | PR | FM |
| J48 | 0.757 | 0.779 | 0.779 | 0.78 | 0.78 | 0.779 |
| J48 binary tree | 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 |
| Naïve 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 |  |  |  |  |  |  |

Table 4. The best learning algorithm result comparison

With the reference to Table 4, the best result of each learning algorithm is introduce. In the best result of decision tree have gained by the implementation of MT applied with CLBP and CFS, while decision tree with 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 is 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 Naïve 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 reference to the implementation of LF applied with RLBP and CFS. Lastly, the multilayer find out by LF applied with RLBP and CFS for AUC, LT applied with Gabor and CFS for SP and PR value, and others by LF applied with RLBP and CFS or LT applied with Gabor and CFS.

1. **Conclusion**
2. **Acknowledgement**
3. **Reference**