Early Detection of Osteoarthritis Stage by Applying Classification Technique to Human Joint imagery

Sophal Chan

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology
Prince of Songkla University
2019

Thesis Title Early Detection of Osteoarthritis Stage by Applying Classification Technique to Human Joint imagery

Author Mr. Sophal Chan

Major Program Information Technology

Major Advisor

……………………………………

(Dr. Kwankamon Dittakan)

Examining Committee:

……………………..…Chairperson

(Dr. Korawit Prutsachainimmit)

………………………Committee

(Dr. Sanparith Marukatat)

………………………Committee

(Dr. Kwankamon Dittakan)

The Graduate School, Prince of Songkla University, has approved this thesis as partial fulfillment of the requirements for the Master of Science Degree in Information Technology

…………………………………………

(Prof. Dr.Damrongsak Faroongsarng)

Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate’s own investigations. Due acknowledgement has been made of any assistance received.

....Signature
(Dr. Kwankamon Dittakan)
Major Advisor

....Signature
(Mr. Sophal Chan)
Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

....Signature
(Mr. Sophal Chan)
Candidate

Thesis Title Early Detection of Osteoarthritis Stage by Applying Classification Technique to Human Joint imagery
Author Mr. Sophal Chan
Major Program Information Technology
Academic Year 2018

ABSTRACT

Knee Osteoarthritis (OA) is a degenerative joint disease or degenerative arthritis which is the most common chronic condition of joint inflammation coursing various paints such as joint paint, stiffness, swelling, creaking or creating sound, decrease ability to move and bone spur. It is a major cause of disability in older people. The risk of knee OA increases from age 45 and older. Early diagnosis is typically made using X-ray imagery. The work presented in this thesis has proposed the three different mechanisms for knee OA Classification includes: (i) texture based approach, (ii) graph-based approach, and (c) Convolutional Neural Network (CNN) deep learning approach. The fundamental idea is to segment X-ray image so as to obtain the X-ray pixels describing the region of interest (ROI) which were done manually and representing these segmentations using an appropriate representation mechanism to translate an X-ray image into a form that serves to captures key information while remaining compatible with the classification process. By pairing each representation with its OA stage, a classifier can be generated to predict the OA stage according to the nature of a selected representation. The generated classifier can be then used to provide a quick and easy mechanism for labelling the X-ray imagery. As the result of the three different approach, the texture-based approach produce the best result of AUC (Area Under Curve) value of 0.912 in case of OA detection and AUC value of 0.871 in case of OA stage classification. For graph-based approach, the OA detection performed with the AUC value of 0.917 and OA stage classification with AUC value of 0.819 as the best record of graph-based. Finally, the application of CNN produced the best result of OA detection study with AUC value of 0.880 and OA stage classification with AUC value of 0.629. It can be concluded that all the three approach: (i) Texture-Based,

(ii) Graph-Based and (iii) CNN-based are well performed on OA detection, while texture-based and graph-based produced the good result for OA stage classification.

Keywords: Image classification, Image Analysis, Image Processing, Osteoarthritis, X-ray image Analysis, Data Mining, Osteoarthritis Analysis, Osteoarthritis Stage Analysis, Medical Image Processing,

ACKNOWLEDGEMENT

I would first like to thank my thesis advisor Dr. Kwankamon Dittakan of the College of Computing at prince of Songkla University, Phuket campus. The door to Dr. Kwan office was always open whenever I ran into a trouble spot or had a question about my research or writing. She consistently allowed this paper to be my own work, but steered me in the right the direction whenever she thought I needed it.

I would like to give a big thank to all the members of my committee for their patience, sharing knowledge and giving assistant in overcoming numerous obstacles I have been facing through my research. I would like to express my gratitude to the Bangkok hospital (Phuket branch) and Dibuk hospital for collecting and providing the X-ray data set to my research. I would like to express my gratitude to MD. Sirisak Yaisoongnern for his managing process for data set collection and MD. Chaowakon Saehang for giving his valuable knowledge for OA stage classification.

Finally, I must express my very profound gratitude to my family and my friends for encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Sophal Chan

CONTENTS

CONTENTS (cont.)

CONTENTS (cont.)

CONTENTS (cont.)

LIST OF FIGURES

LIST OF FIGURES (cont.)

Figures **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Introduction

"Knowledge Discovery in Data (KDD) is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" (Fayyad, et al.,1996). With respect to the work presented in this thesis data mining is an essential element which concerned with the discovery of the desired hidden information from huge or complex data set. Data mining can be performed on any kind of data, the example includes: a relational database (Han, et al., 2006) where data is stored in a table that has a two dimensional structure with row and column; a data warehouse is an enormous subject-oriented database integrated from multiple sources in a given time period for decision-making process; a transactional database involves day-to-day operation data for example from banking or supermarket transactions; advanced database systems are provided to meet the requirement of new database applications, for example object-oriented and object-relational databases, and specific application-oriented database such as spatial data, time-series data, text data, multimedia data, and World Wide Web. Consequently, data mining includes sub-fields such as image mining, graph mining, and text mining. The work described in this thesis is concerned with image mining.

Image mining is a mechanism for extraction of useful knowledge and correlations within image sets. Large amounts of visual information (in the form of digital images) are generated on a daily basis with respect to many domains such as the remote sensing images and medical images domains. Extracting useful knowledge from within these images presents a significant challenge. Image mining can be applied in other fields such as image analysis and content-based image retrieval.

According to the work presented in this thesis is directed at image classification which is a non-trivial problem, because of the typically complex structure of image data, and is still a very active field of research. With respect to image classification a collection of prelabelled images is taken as input and used to generate (train) a classifier which can then be applied to unseen images. Image classification typically involves the preprocessing of collections of images into a format whereby established classification techniques could be applied. As with many data mining applications the main challenge in the preprocessing of image data is to produce a representation whereby no relevant information is lost while at the same time ensuring that the end result is accurate enough to allow for the application of effective data mining.

With respect to the research motivation, Osteoarthritis (OA) is the well-known human joint disease. It had affected 10% of Thai population in 2014. When people are affected by OA, it is difficult to recover back as normal. Thus, the most typical way is medical imaging for OA early detection in order to prevent to the serious condition. One of the most common OA diseases is knee OA. Knee OA have affected million Asian people, from the Community Oriented Program for the Control of Rheumatic Diseases (COPCORD) studies and (Fransen, et al., 2011 and Cho, et al., 2011) studies illustrated that in Asian aging society Knee OA have affected in range from 38.10% to 50.00%.

The fundamental of this thesis presents that the collection knee Osteoarthritis (OA) X-ray images can be applied to classification technique for classifying the stages of knee OA. The research illustrated is thus directed at mechanisms for the construction of a classification model that can predict stages of knee OA according to the nature of screening X-ray imagery.

Classification is considered as the final process of the thesis methodology. This process is the mechanism of generating a classifier that could be used to describe data classes. The classifier is derived using label training data. Classification has been widely applied in many areas including medical diagnosis, customer segmentation, weather detection, fraud detection, and weather prediction (Anitha, et al., 2014; Fesharaki, et al., 2012 and Zu, et al., 2008). The classification can be applied to predict discrete and/or unordered class label. The accuracy of predictive model is defined by using it to pre-label test data.

The remain of this chapter is organised as follows. Section 1.2 presents the research objectives and associated research issues and challenges. The research methodology used to describe the research challenges, including the “criteria for success”, is pictured in Section 1.3. In Section 1.4 describes the contribution of the research work, and the published work to date arising from the research, is illustrated in Section 1.5. The overview of the rest of this thesis is presented in Section 1.6. Finally, the summary of this chapter is described in Section 1.7.

1.2 Thesis Objective

From the forgoing the research concerning with the investigation, realization and evaluation that can be create classification models for the purposed of two studies include: (i) knee OA detection study and (ii) knee OA stage classification study. The thesis objective in this work is encapsulated by the following research question:

Can the knee OA and the stage knee OA can be predicted by applying classification technique and deep learning model to human joint X-ray imagery?

The dedication of this research question envelop a number of research challenge. There are the articulated below are listed in the form of a series of subsidiary research questions:

1) How to obtain the Region of Interest (ROI) from X-ray images?

2) What is the information that should be extracted from the identified sub-images and how can this information best be extracted?

3) Once the desired information has been extracted what is the best way of representing these images so as to support the effective generation and usage of classifiers?

4) What are the most appropriate classification techniques for stage of OA detection from given image in the context of different information?

5) What is the most appropriate value of support threshold for stages of OA detection from given image in the context of different information?

6) Is the deep learning-based work powerfully to the knee OA stages detection?

The thesis work out to provide solutions to the above questions.

1.3 Research Methodology

To act as a focus for the work a set of medical X-ray images was collected from two local hospitals: (i) Dibuk hospital (Phuket branch) and Bangkok hospital (Phuket and Suratthani branch), these collection of these X-ray images were used in this thesis work. As noted in Section 1.2 Thesis objectives, there are two studies had been evaluated: (i) Knee OA detection and (ii) knee-OA Stage classification. With respect to knee OA detection, the images were categorised into two different classes: (i) knee with OA and (ii) normal control (knee without OA) by domain expert from Bangkok hospital; the example of knee with OA and without OA are presented in Figure 1.1.

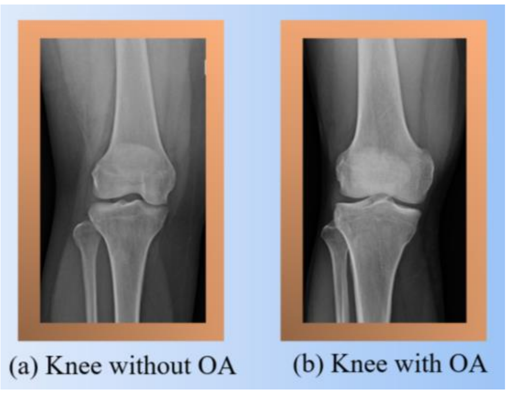


Figure 1.1 Normal and knee OA condition

Figure 1.1 (a) illustrates the example of X-ray image from the patient without OA and Figure 1.1 (b) presents example of X-ray image from the patient with OA.

With respect to knee-OA stage classification, the image were categories into five different groups based on Kellgren and Lawrence (Kellgren, et al., 1963 and Kellgren and Lawrence, 1957) system and the image was graded by domain expert from Thungsong hospital, Nakhon Si Tham Rat province. The five groups of OA stages are separated from *stage* 0 till *stage* 4; the stages of knee-OA are illustrated in Figure 1.2:

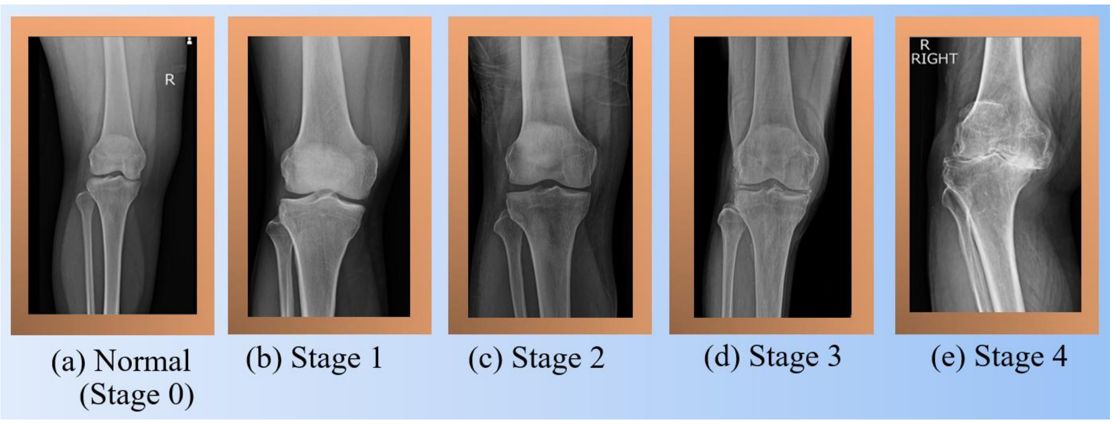


Figure 1.2 The Stage of Knee OA

From Figure 1.2 it can be seen that Figure 1.2 (a) presented the *stage 0* of knee-OA stage or normal control; Figure 1.2 (b) illustrated the *stage* 1 of knee-OA stage or "Doubtful stage"; Figure 1.2 (c) presented the *stage* 2 or "Mild stage" of knee-OA stage; Figure 1.2 (d) illustrated the *stage* 3 or the moderate grade of knee-OA stage. Lastly, the Figure 1.2 (e) presented the *Stage* 4 which is the serious condition of knee-OA stage or severe grade. Prior the classification could be commenced the Region of Interest (ROI) or the sub-image from an X-ray image must be identified.

The first step in the methodology of this work was to segment an X-ray image into regions. In this process, the manually segmented was applied due to the great challenge of knee image in term of texture and shape of object.

The second step was to investigate algorithms for representing ROI of knee image in such a way that: (i) compatibility with classification model generation techniques was obtained and (ii) The minimum losing of information. A review of the existing related work and literature involvement with image classification suggested three main groups of representing technique: (i) texture-based and (ii) graph-based. For the learning the information (learning feature) of object also presented in the work, with the respect to the literature, the Convolutional Neural Network deep learning-based can be applied with the image representation techniques. Thus, there are thee approaches for the study: (i) texture-based, (ii) graph-based, and (iii) deep learning-based.

In the context of texture-based and graph-based study, the third step (the last step) in the proposed methodology was to consider a variety of classifier generation. In the literature review, there are a big number of these with no clear "best" model generators. To define the most appropriate the idea was to conduct a significant amount of evaluation combining each of the proposed representation with a number of different generators. The criteria for success of this thesis work was prediction accuracy, comparison predicted stage of knee OA with known the knee OA stages. With reference to the evaluation pictured later in this thesis, in term of the classification models, results were consider on the subject of Area Under Receiver Operating Curve (AUC), Accuracy, Specificity, Sensitivity, Precision, and the F-measure; of which AUC was considered as the most significant. Later in this work results are illustrated in the form of table, with the later focusing only on the AUC value from the tables.

1.4 Contributions

The contributions of the research study illustrated in this this thesis can be briefly described as follows:

1) A knee sub-image (ROI) representation founded on the concept of "texture" analysis. More specifically applying of Local Binary Patterns (LBPs), as before a feature vector format was build.

2) A knee sub-image representation founded on the concept of "graph-based" by applying the quadtree hierarchical decomposition together with frequent subgraph mining for reducing the feature dimensionality. The identifier frequent

subgraph were set to a feature vector format, one vector per ROI, suited for input into a learning classifier.

3) An approach of deep learning-based for classification without any performance from manually feature extraction.

4) An analysis of a sequence of the proposed sub-image (ROI) so as to identify the most appropriate ROI in term of knee OA detection from X-ray images.

5) An analysis of a sequence of the proposed ROI image representation algorithm so as to select the most appropriate in term of knee OA detection from X-ray images.

6) An analysis of a sequence of feature selection algorithm so as to select the most appropriate in term of knee OA detection from X-ray images.

An analysis of a sequence of classifier generation algorithm so as to select the most appropriate in term of knee OA detection from X-ray images.

1.5 Publications

A number of research paper have appeared from the work illustrated in this thesis. There are three groups of the research papers were categorized in this thesis includes: (i) the accepted or published papers, (ii) the submitted papers, and (iii) the papers in the preparation process. These are itemised below, in each case a brief information is given and a reference to where the material feature in this thesis.

1) The Accepted Paper

Sophal Chan and Kwankamon Dittakan: *Osteoarthritis Stages Classification to Human Joint Imagery using Texture Analysis: A Comparative Study on Ten Texture Descriptors*. RTIP2R Conf. 2018.

This paper was described the study of knee-OA stage classification. The major ideas presented in this paper was separated into three groups: (i) the ROI that produced the best performance of classification in case of texture approach, (ii) the image representation algorithms or texture descriptor techniques that made the best result of classification, and (iii) the classifier generation methods that produced the top performance of knee-OA stage classification. The presented evaluation indicated that the knee-OA stage can be classified efficiency by using the texture analysis. The work summaries some of the material illustrated in Chapter 4 where the detail of the proposed texture-based representation is discussed.

2) The Submitted Paper

Sophal Chan, Kwankamon Dittakan, and Matias Garcia-Constantino: *A Comparative Study of Texture Analysis Techniques for Osteoarthritis Classification Using Knee X-ray Imagery.* Submitted to Journal of Digital Image 2018.

This paper was described the study of OA detection. The main ideas illustrated in this paper was separated into four groups: (i) the ROI segmentation that produced the best performance of OA detection in case of texture approach, (ii) the image representation algorithms or texture descriptor techniques that made the best result study, (iii) the best feature selection technique for OA detection study, and (iii) the classifier generation methods that produced the top performance of OA detection study. The presented evaluation indicated that the OA detection can be detected by using the texture analysis of the proposed approach. The work summaries some of the material illustrated in Chapter 4 where the detail of the proposed texture-based representation is discussed.

3) The Preparation Paper

Sophal Chan and Kwankamon Dittakan: *Osteoarthritis Stages Classification to Human Joint Imagery using Quadtree Analysis: A graph-based Study on Knee Medical X-ray Images*.

Sophal Chan and Kwankamon Dittakan: Osteoarthritis Stages Classification to Human Joint Imagery using Convolutional Neural Network: A deep learning-based for classification on Knee Medical X-ray Images.

1.6 Thesis Organisation

The arrangement of the rest of this thesis is as follow. Chapter 2 provides an extensive literature review of image analysis in term of knee OA detection and the previous work concerning the technologies that feature in this thesis. Chapter 3 reported the way to identify ROIs in order to apply further in the next chapter. For Chapter 4 is described the OA detection and the classification of knee-OA stage study by applying texture-based approach. Graph-based approach of using the quadtree decomposition was applied to detect OA and classify knee OA stages is presented in Chapter 5. The application of convolutional neural network for stage of knee OA classification is discussed in Chapter 6. Finally, in Chapter 7 the thesis conclusion include summary, presentation and discussion of the main findings in term of the research question and sub-questions defined above and some direction for future work.

1.7 Summary

This chapter has given the necessary context and background for the research work presented in this thesis. Specifically, the motivation for the research and the research objectives have been illustrated. In the following chapter (Chapter 2) presents a literature review of the related previous work. The arrangement of the rest of this thesis is as follow.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

2.1 Introduction

A review of the background and previous work in the context to this research in thesis is presented in this chapter. The content of this chapter is organised as follows Section 2.2 Medical Image, a brief history of medical image technology and the application of medical image in human debases diagnose, typically Knee-OA is presented. Section 2.3 Image Processing, the basic idea of image processing technology is described. There are three image processing techniques: (i) Image Segmentation, (ii) Image Enhancement and (iii) Feature Extraction/Image Representation. The brief detail of each technique is presented in Sub-section 2.3.1, Sub-section 2.3.2 and Sub-section 2.3.3, respectively. The summary information of image classification is suggested in Section 2.4. The detail of the section is described as follows: (i) the fundamental idea of data mining (ii) the detail of image classification and (iii) the classification learning methods are presented Sub-section 2.4.1 to 2.4.3, respectively. The feature selection techniques which apply to select the feature vector are described in Section 2.5. Section 2.6 points out of the brief information of Convolutional Neural Network (CNN) applying in medical imaging. The evaluation measurement methods which is used to evaluate the classification result is discussed in Section 2.7, in this section there are six measure parameters including: (i) Accuracy (AC), (ii) Sensitivity (SN), (iii) Specificity (SP), (iv) Precision (PR), (v) F- Measure (FM), and (iv) Area under the ROC Curve (AUC). In case of AC, SN, SP, PR, and FM can be calculated from Confusion Matrix which is presents in Sub-section 2.7.1. While AUC can calculates from the Receiver operating characteristic (ROC) curve which is presented in Sub-section 2.7.2. The comparison of some related works are presented in Section 2.8. Finally, the summary of this chapter is described in section 2.9.

2.2 Medical Image

Medical image has spread widely and has been an interdisciplinary to many research fields, for example, computer science, computer vision, engineering, statistic, biology, and medicine. The technology in computer-aided diagnose processing has become an important for clinical healthcare. The medical image is used to diagnose inside human structure such as bone, brain and renal system. There are various types of medical image with respect to image capturing mechanisms including (i) X-ray (radiography), (ii) Computed Tomography (CT), (iii) Magnetic Resonance Imaging (MRI), (iv) Ultrasound, (v) Positron Emission Testing (PET) and (vi) Single Photon Emission Computed Tomography (SPECT). With respect to the work presented in this research, to act as a focus for the work X-ray image is considered. However, there is three medical image capture mechanisms are presented in this section including (i) X-ray, (ii) CT and (iii) MRI.

2.2.1 X-ray Image

German physicist Wilhem Rontgen discovered X-ray image in 1895 (Thescientist, 2018). The first medical image of X-ray image was taken from his wife's left hand. Wilhelm Rontgen took this radiograph of his wife's left hand on 22 December 1895, shortly after his discovery of X-rays. The image of Wilhelm's wife's hand is shown in Figure 2.1.

Figure 2.1 The image of Wilhelm wife left hand (Thescientist, 2018)

X-ray is a form of electromagnetic radiation which can pass through the human body with producing the internal structure image, the resulting of X-ray is an image, and this image is called a radiographic, also known as “X-Ray” or “plain film”.

Figure 2.2 An example of X-ray image with different radiographic (Lisle, 2012)

There are five categories depending on radiographic densities in X-ray image. The Figure 2.2 illustrates the difference of X-ray radiographic density indicating by the number including (i) number 1 is air (the darkest area), (ii) number 2 presents fat (the dark grey), (iii) number 3 is the soft tissue (grey), (iv) number 4 illustrates bone (bright grey) and (v) number 5 is the contrast material (Lisle, 2012).

X-ray images have been used to diagnose Knee-OA which presented in many works (Esteva, et al., 2017; Pandey, et al., 2016; Gornale, et al., 2015; Shamir, et al., 2009a, 2009b, 2010; Shamir, 2011 and Herman, et al., 2015). There are two main objectives for Knee-OA research: (i) X-ray image was used to detect OA and non OA (Pandey, et al., 2016 and Gornale, et al., 2015), while the second one is more advanced than the first one because the X-ray image has been used to detect the OA stages as reported in (Shamir, et al., 2009a, 2009b, 2010; Shamir, 2011 and Herman, et al., 2015).

On the purpose of the detection of the OA stage, the Kellgren-Lawrence (KL) system (Kellgren, et al., 1963 and Kellgren and Lawrence, 1957) is considered. The KL standard is a five stages indicating using *stage* 0 to *stage* 4. '*Stage* 0' represents no OA symptom while '*Stage* 4' represents the most serious condition of radiographic disease. The summary of the OA stage is shown in Table 2.1. There are five stage of OA has shown in the table which rank from zero to fourth stage. The *stage* 0 refers to the normal case, within this stage the sign of OA cannot detect because there is no sign of OA has happened. The second stage of OA is the doubtful stage sometime called *stage* 1 of knee OA; in this stage there is a little sign of OA has appeared. In other words, in the *stage* 1 of OA there is a sign to show very minor bone spur growth. The next stage is the *stage* 2 are considered as the minimal stage, the knee joints in this stage will reveal greater bone spur growth, but the cartilage is usually still at a healthy size, i.e. the space between the bones is normal, and the bones are not rubbing or scraping one another. At this stage, synovial fluid is also typically still present at sufficient levels for normal joint motion. However, this is the stage where people may first begin experiencing symptoms; pain after a long day of walking or running, greater stiffness in the joint when it's not used for several hours or tenderness when kneeling or bending. *Stage* 3 of OA is considered as "moderate" stage. In this stage, the cartilage between bones shows obvious damage and the space between the bones begins to narrow. People with *stage* 3 OA of the knee are likely to experience frequent pain when walking, running, bending, or kneeling. They also may experience joint stiffness after sitting for long periods of time or when waking up in the morning. Joint swelling may be presented after extended periods of motion. *Stage* 4 OA is considered as "severe." People in *stage* 4 OA of the knee experience great pain and discomfort when they walk or move the joint. This is because the joint space between bones is dramatically reduced; the cartilage is almost completely gone, leaving the joint stiff and possibly immobile. The synovial fluid is decreased dramatically, and it no longer helps reduce the friction among the moving parts of a joint.

Table 2.1 The summary of the knee OA stages

Knee OA Stage	***Stage* 0 (Normal)**	***Stage* 1 (Doubtful)**	***Stage* 2 (Minimal)**	***Stage* 3 (Moderate)**	***Stage* 4 (Severe)**
Properties	No sign of OA	Show very minor bone spur growth	Reveal greater bone spur growth	Cartilage was damage and narrowing of joint space	Great pain and discomfort when walking

Figure 2.3 shows the X-ray image of each OA levels Figure 2.3(a) presents *stage* 0 OA or normal, Figure 3 (b) illustrates *stage* 1 OA (doubtful), Figure 3(c) shows *stage* 2 OA the minimal level and Figure 3 (d) presents *stage* 3 moderate level.

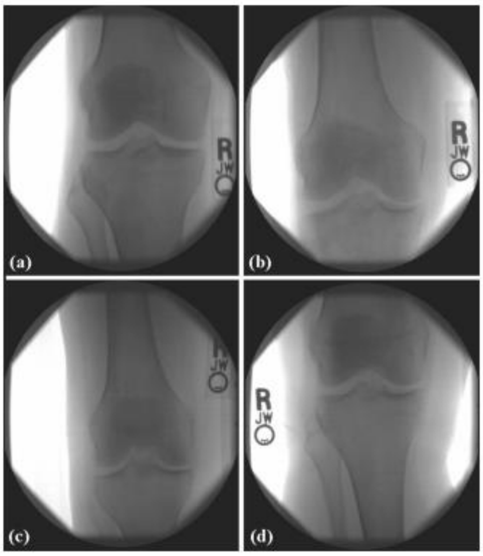


Figure 2.3 X-ray image of four different KL grade (Shamir, et al., 2009)

The work of OA stage detection using X-ray offers the benefit over the detection the OA and non-OA because knowing the level of OA can make orthopaedic doctor easy to prevent and slow down the OA speed to the OA patient. The most important is the cost because the price of X-ray image is inexpansive compare to the others.

2.2.2 Computed Tomography

Computed Tomography (CT) is one of the well-known medical imaging for disease diagnosis. A CT-scan uses a collection of X-ray images taken from different angles to produce cross-sectional image using the computer processing. Thus the structures inside of the body can be presented without cutting. CT provides the clear image so the medical doctor can be used to analyse or diagnose the disease.

2.2.3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) uses the magnetic properties of spinning hydrogen atoms to make the internal structure image. It makes the powerful image of human bone structure and create a very clear image. With respect to OA detection, the MRI images were also applied as presented in (Kubakaddi, et al., 2013; Sargunar and Sukanesh, 2009; Bindu, 2009; and Senthilkumaran and Vaithegi, 2016).

The works show the benefit of OA analysis with short time consuming and produce a good result. The example of MRI image of human knee is shown in Figure 2.4.

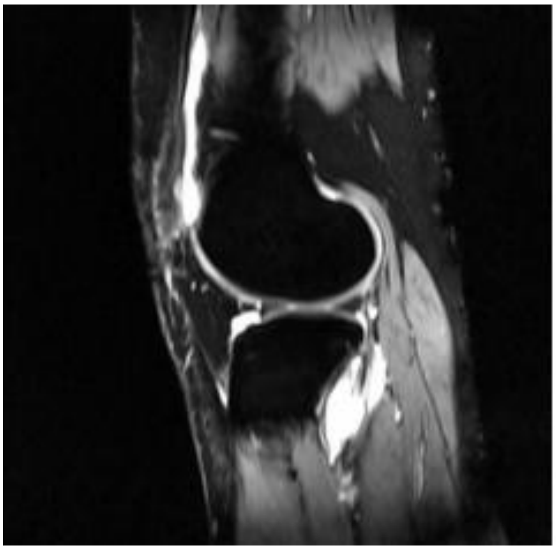


Figure 2.4 Human knee image using MRI (Kubakaddi, et al., 2013)

To summarise, OA detection is the challenge task in medical and computer science. The image processing is applied to OA detection using medical image is the best solution as well as to analyse the OA stage. With the respect to this research to act as a focus in Thailand, so the medical X-ray image is considered with low cost of processing. With X-ray image apply with image processing to diagnose OA stage, the image processing technique is presented in the next section.

2.3 Image Processing

The fundamental concepts of image processing is discussed in this section. Image processing is a challenge area in computer science. The aims of this method is to analyse and manipulate the digital image. In addition image processing has been applied to improve the image quality. There are main three key words in image processing that can be listed as follows:

(i) Image processing (input as image and output as image)

(ii) Image analysis (input as image and output as the measurement)

(iii) Image understanding (input as image and output as the high-level description)

An image is a set of array, a matrix, a square of pixel which arrange in columns and rows. In 1920s, the first image processing was discovered and it was called Bartlane cable picture transmission system. Nowadays, image processing has been applied in various applications such as education, engineering and medical. To get the deep meaning of image processing, the definition of an image should be known clearly first. An image is considered of two real variable, for example I(x,y), with I as the amplitude of the image at the real coordinate position (x,y). Sub-image is the sub set of image, sometime it considers as the region of interest (ROI). The smallest element of image is pixel. Image processing generally separate into two main categories: (i) digital image processing and (ii) medical image processing (Maintz, 2005).

In general the medical image processing provides the image in white and black colour, or greyscale image in technical word. An 8-bits greyscale image has which picture element has an assigned intensity that ranges from 0 to 255. Figure 2.5 shows each pixel of greyscale image with the value from 0(black) to 255 (white) because 8 bit equal to 256 greyscales.

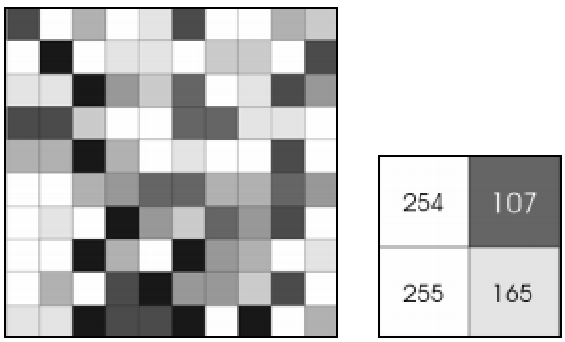


Figure 2.5 Each pixel has a value from 0 (black) to 255 (white) (Spacetelescope, 2016)

There are three image processing processes including: (i) image segmentation, (ii) image enhancement and (iii) image representation/image feature extraction. The detail of each process is discussed in the Sub-section 2.3.1 to 2.3.3 respectively.

2.3.1 Image Segmentation

Image segmentation is one of the most important processes of the image processing. The main objective of image segmentation is to divide the given image in to salient image regions, meaningful region, and homogeneous with specific cluster pixel. In addition, the image segmentation is applied to various application domains includes: object recognition, occlusion boundary estimation within motion or stereo systems, image compression, image editing, or image database look-up. A good segmentation of an image can be defined by: (i) pixel in the same region or categories has similar or equal pixel colour/intensity and (ii) the neighbouring pixel in different categories has different values or dissimilar values.

Although image segmentation is one of the main process in image processing, but it has many algorithms to apply with based on the different kind of work. Some of the existing image segmentation methods that are widely used are presented here: (i) threshold-based segmentation, (ii) region-based segmentation and (iii) edge-based segmentation. The brief detail for each method is discussed as follow:

Threshold segmentation

Threshold-based segmentation is one the existing methods for image segmentation. The separation of bright and dark region is the main objective of threshold segmentation. In general, thresholding is applied to create the binary image. For all pixel of greyscale image is calculated by comparing each pixel with threshold (T). If the intensity values is below than threshold the pixel which make from grey-level ones by turning all pixels that below threshold become zero (0) and if pixel is above threshold the pixel become one (1). For example: if I(x, y) is a threshold version of i(x,y) at some global threshold T, as Equation 1.1 bellows:

$$I(x,y)=\begin{cases} 1 \text{ if } i(x,y) \geq T \\ 0 \quad \text{otherwise} \end{cases} \qquad 2.1$$

Threshold-based segmentation has a drawback which is only the pixel value is used to consider on other hand no relationship between the pixels is examined. Therefore, there is no any guarantee that the pixels identified by the thresholding process are contiguous.

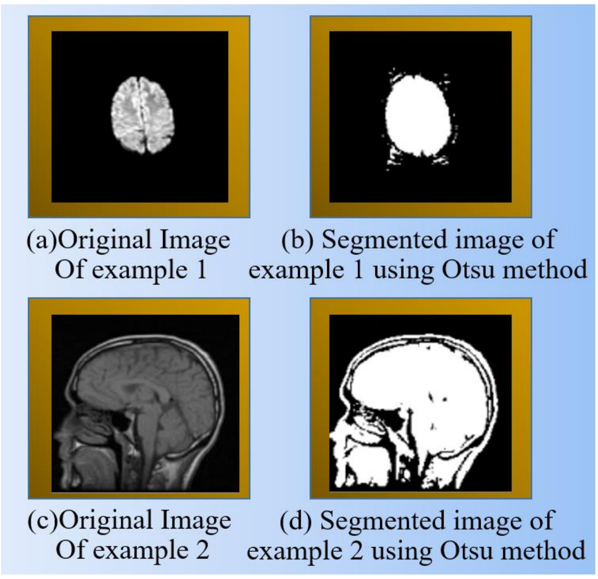


Figure 2.6 The human brain scan using Otsu method (Bindu, 2009)

There are many applications using threshold mechanisms were proposed for medical image segmentation. In work (Bindu, 2009) the implementation of Otsu method to improve medical image segmentation. Otsu method is a method to reduce the greyscale to a binary image, the result shown that the method performs better than other thresholding methods with a good binary images. Otsu threshold method is a very good and efficient method to threshold the greyscale image. However, the drawback of this method is high complexity of the computation. Figure 2.6 (a) and Figure 2.6 (c) shows the original of human brain scan, while Figure 2.6 (b) and Figure 2.6 (d) presents the corresponding segmented image using Otsu method.

Local adaptive thresholding also is applied in (Senthilkumaran and Vaithegi, 2016). The work reported the implementation of local adaptive threshold technique to removes background using local mean and standard deviation. The comparative study between Niblack and Sauvola local thresholding had been evaluated. The results shown that the Niblack local threshold algorithm reduced the background noise better than Sauvola local threshold algorithm. Figure 2.7 shows the difference of five sample sub-images between two different methods: Niblack technique and Sauvola technique.

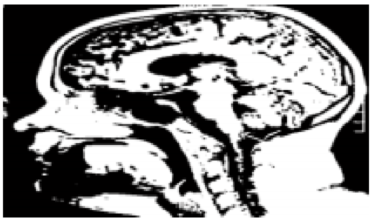
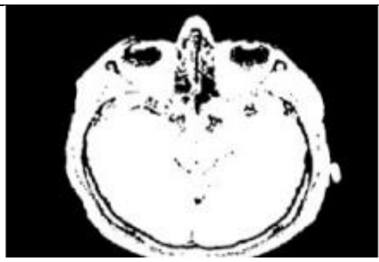
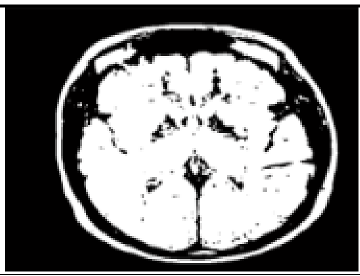
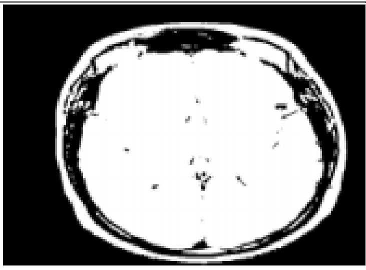
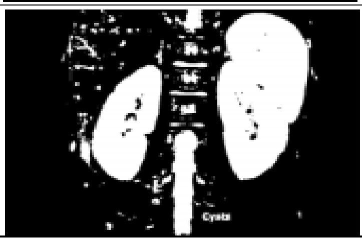
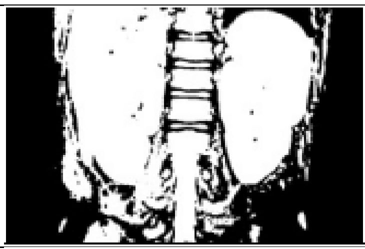
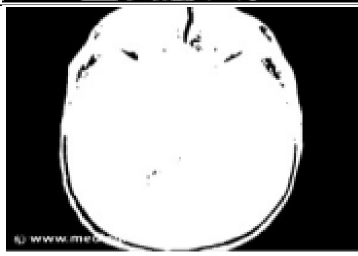
Images	Using Niblack Technique	Using Sauvola Technique
Image 1		
Image 2		
Image 3		
Image 4		
Image 5		

Figure 2. 7 Example of The comparison of five medical images by using Niblack and Sauvola Technique in (Senthilkumaran and Vaithegi, 2016)

Region based segmentation

Region-based segmentation is one of image segmentation techniques. The main objective is to find coherent regions in the image. The coherent region contains pixels that has same or some similar property. The transitivity of the similarity relationship in the image is considered. Region-based segmentation offers the advantages on (i) fast processing and (ii) more efficient than edge-based and threshold-based method. However the region-based also has the drawback concerning it probably grow further away from global pixel because it drifts as one. There are many existing region-based segmentation algorithms for example watershed segmentation, flooding-based watershed segmentation, marker-controlled watershed segmentation and inter-pixel watershed segmentation. The implementation of watershed segmentation in the context of medical image with texture-based region merging was proposed in (Ng, et al., 2008). The work in (Ng, et al., 2008) reported the texture-based region merging was applied to MRI images in order to improve the segmentation efficiency. The comparison of images using watershed segmentation with/without text-based region merging and with texture-based region-based region merging is illustrated in Figure 2.8.

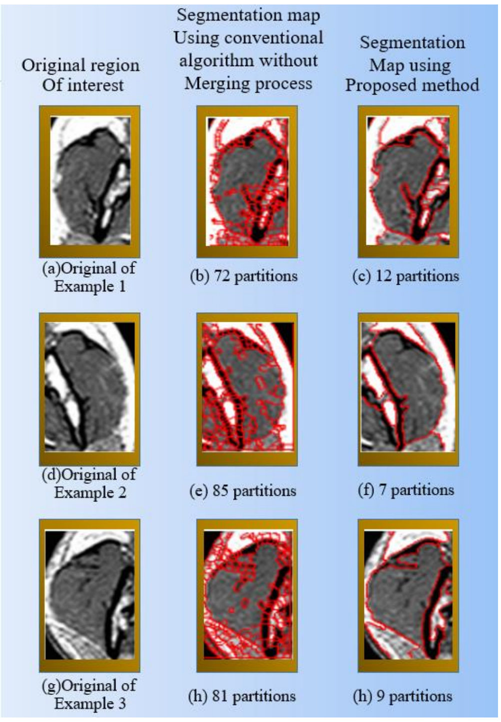


Figure 2.8 The comparison of image using watershed segmentation with/without texture-based (Ng, et al., 2008)

Edge based segmentation

In edge-based segmentation mechanism focuses on the edge in an image. The edge corresponds to singularities in the images. The edge in image generally represents as the shape of objects in the scene. The purpose of edge-based segmentation is to extract the edge or line in the image with good orientation. There are some existing edge operators for example Gradient operator, Prewitt operator, Sobel operator, Compass operators and Laplacian operator. Figure 2.9 shows the Prewitt and Sobel operator.

$$
\begin{array}{lcc}
 & \text{(a) Prewitt operator} & \text{(b) Sobel operator} \\
\text{Verticel} & \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} & \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \\
\text{Horizontal} & \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} & \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}
\end{array}
$$

Figure 2.9 Prewitt and Sobel operator for edge detection

With the respect to the edge detection in the context of medical image segmentation, it can be found in (Bandyopadhyay, 2011). The edge detection was applied to the brain scan images to detect the tumour regions based on the gradient magnitude information. The example image after applied the edge detection is shown in Figure 2.10.

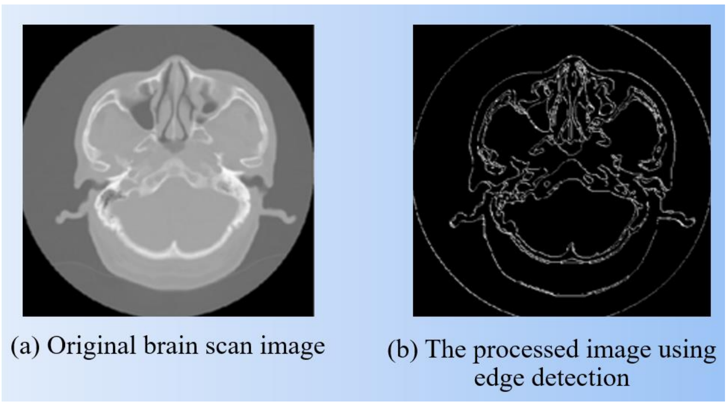


Figure 2.10 The example of brain scan image using edge detection (Bandyopadhyay, 2011)

From Figure 2.10 (a) shows the original brain scan while Figure 2.10 (b) illustrates the brain scan image after the edge detection was applied.

2.3.2 Image Enhancement

Image enhancement is one of the most important and complex techniques in image processing. The fundamental idea of image enhancement mechanism is to improve visual appearance of an image, or to present a better transform representation of the image. The main challenge in image enhancement is quantifying the criterion for enhancement.

There are some various types of image such as personal images, medical images, satellite images and aerial images. The photographers/users/viewers probably suffer from some image quality problems such as poor contrast and noise. Therefore image enhancement plays the main role to enhance the contrast and remove the noise to increase image quality. During image enhancement process an input is an original image and the output also an image with better than the input image by changing the pixel intensity. Furthermore, image enhancement works as the important roles in many kinds of image for example hyperspectral image processing, remote sensing, high definition television (HDTV), industrial X-ray image processing, microscopic imaging, other image/video processing applications and medical image processing (Zhang, et al., 2011).

In addition image enhancement technique can be used to increase dynamic range of the chosen features in of an image such as point operation, spatial operation, transform operation and pseudo colouring. There are some existing methods of image enhancement for example: filtering with morphological operators, histogram equalization, noise remove using a Wiener filter, linear contrast adjustment, median filtering, unsharp mask filtering, contrast-limited adaptive histogram equalization (CLAHE) and decorrelation stretch.

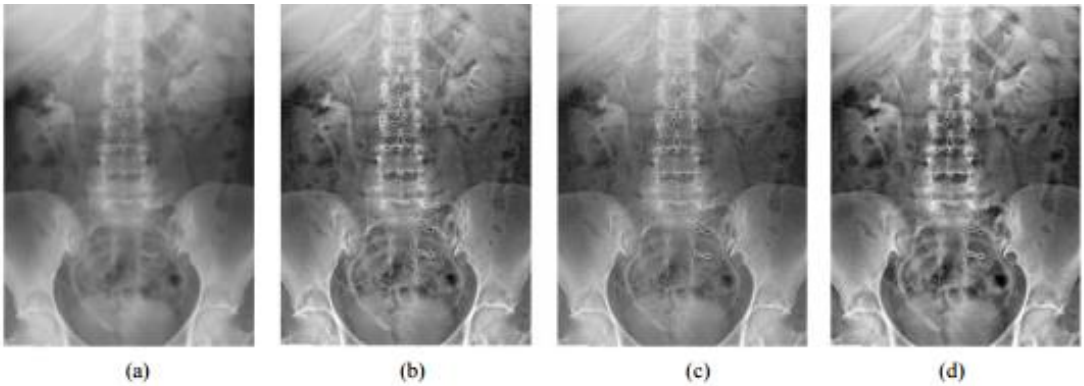


Figure 2.11 The example of X-ray image of renal system using morphological transform (Firoz, et al., 2016)

In the context of medical image, image enhancement also applied in many applications. Morphological transform was applied to medical images (Firoz, et al., 2016). The work reported on using a disk shaped mask is used in Top-Hat and Bottom-Hat transform in order to remove noise and increase the contrast quantity in medical images. Figure 2.11 shows the morphological transform was used in X-ray of renal system. Figure 2.11 (a) presents the original image. Figure 2.11 (b) illustrates optimum enhanced image. While Figure 2.11 (c) and Figure 2.11 (d) shows the transformed image with different radius using Carrier-to-noise ratio (CNR) and Peak signal-to-noise ratio (PSNR), respectively.

2.3.3 Feature Extraction or Image Representation

The fundamental information of feature extraction or image representation is presented in this sub-section. In image processing, feature extraction is the main rule for measuring the data and builds derived values (features) in order to find a good data representation. The objective of feature extraction in digital image is to reducing the amount of resources which required to describe a large set of data. Feature extraction is directed relation to dimensional reduction; dimensional reduction is the process of reducing the number of random variables under consideration (Roweis and Saul, 2000). With the respect to dimensional reduction there are some existing techniques including: Independent component analysis, isomap, principal component analysis (PCA), kernel principal component analysis, latent semantic analysis, partial least squares, multifactor dimensionality reduction, nonlinear dimensionality reduction, multilinear principal component analysis, multilinear subspace learning, Semidefinite embedding, autoencoder and deep feature synthesis.

However in the context of image representation or feature extraction in digital images can be done using various using image properties. There are three main image properties including: (i) colour, (ii) texture and (iii) shape. The detail of each image feature extraction technique is described as follows:

Colour analysis

Colour is simplest features in digital image. It makes an image is more colourful and more interesting to human view (Alejandro and Perciano, 2013). The colour analysis for image representation is a technique that can use widely with RGB image. RGB image is one type of digital image which contain RGB color system; RGB system refers to the system that representation in Red, Green and Blue through computer display view. The level of red, green and blue can range from 0 to 100 percent of full intensity. For each level of red, green and blue is represented by the range of decimal numbers from 0 to 255 (256 levels for each colour), or to be a binary is 00000000 or 11111111. RGB also know as true colour image. With respect to digital image which store as an m-by-n-by-3 (*m x n x 3*) data array that defines red, green, and blue colour components for each individual pixel. The purpose of colour image analysis is to define or calculate the percent of each colour has contained in an image.

Texture analysis

The texture analysis concerns with description of characteristic image properties by textural features. Texture analysis (Gonzalez, et al., 2004 and Zhang and Tan, 2002) is an important approach for describing the region. Texture can be divided into three categories: (i) statistical methods (a collection of statistic that used to described of texture), (ii) structural methods (texture is viewed as consisting of many texels arranged according to some placement rules) and (iii) model based methods (used probability model to modelled the image textures). A number of existing texture analysis models has been proposed including autoregressive model, Gausian Markov random field, Gibbs random fields, Wold model, wavelet model, multichannel Gabor model and steerable pyramid.

Texture analysis is also implemented in the context of medical image. The example is found in (Gornale, et al., 2015) the long range, standard deviation and greyscale entropy texture analysis techniques were applied to X-ray images to detect OA and non-OA, the Haralick feature was calculated in order to measure the texture image in term of contrast, correlation, sum of square, sum of average and homogeneity.

In work (Sargunar and Sukanesh, 2009), Frequency Filtering texture analysis was applied using Fourier transform to calculate filter data as the Equation 2.2

$$V\tilde{D} = F * VD \tag{2.2}$$

where

$V\tilde{D}$ is the filter data from the convolution theorem.

F is the filter in the spatial domain.

The examples various filters are illustrated in Figure 12 including: (i) Low pass filter (Figure 2.12 (a)), (ii) High pass filter (Figure 2.12 (b)) and (iii) Band pass filter (Figure 2.12 (c)). The results after applied given filters to human knee MRI image are presented in Figure 2.12 (d), Figure 2.12 (e) and Figure 2.12 (f), respectively.

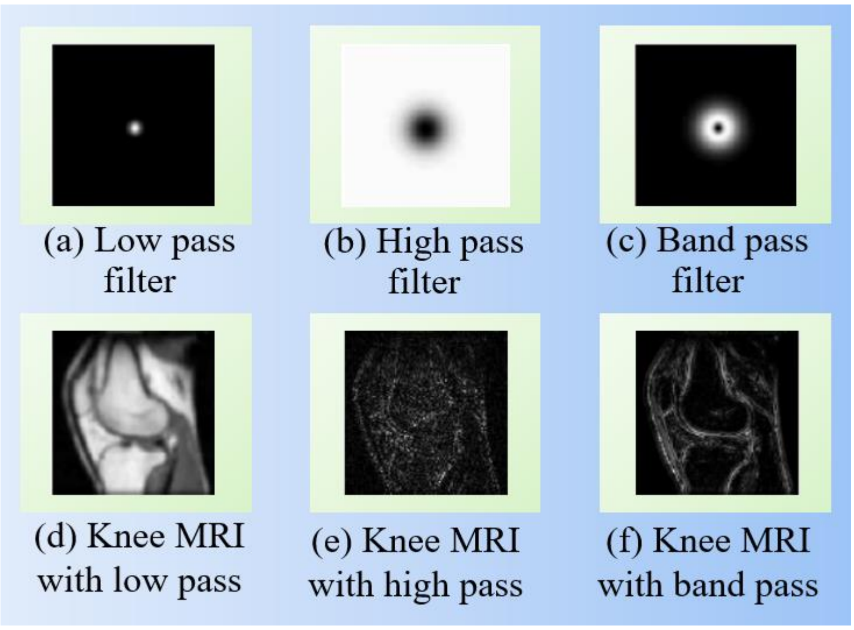


Figure 2.12 The example of various filters applied to MRI of human

Texture analysis

Shape analysis is directed to study the edge of shape feature in image. Shape refers to all the geometrical information that remains when location, scale and rotational effects are filtered out from an object (Dryden and Mardia, 1998). Shape is any connection of point in the image. The example of the existing shape analysis mechanisms including: point distribution, active shape, active appearance models, Fourier Snakes, active contours and parametrically-deformable models. The shape analysis was applied in the context of medical image in many literatures. With respect to the work in this research is directed to Knee OA. There are some examples found in (Maintz, 2005; Kubakaddi, et al., 2013; and Wagaj and Patil, 2015).

In (Maintz, 2005) the shape analysis was used to calculate the thickness between the bone and joint space in knee. Thickness was calculated and found that in the range of 1.69 mm to 2.55 mm was non-OA. In contrast if the thickness is less than 1.69 mm then the symptom of OA was possible. Figure 13 shows the cropping image and applies with binary operation to make edge detection in order to detect the boundary between the bone and joint space.

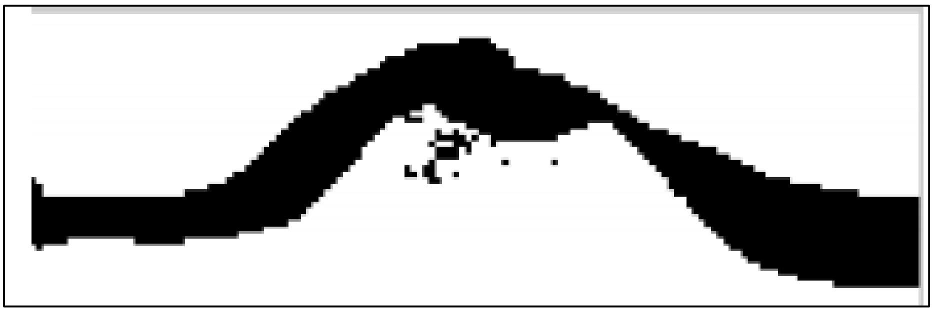


Figure 2.13 Cropping the image and apply in binary operation (Maintz, 2005)

With the respect to the work presented in (Wagaj and Patil, 2015), the shape analysis was applied to calculate the area of cartilage. The initial step was the pixel-based segmentation is used to segment the cartilage. The texture filtering was then applied to calculate the area of cartilage by the number of pixel as presented in Figure 2.14.

Figure 2.14 MRI image of knee after the texture filtering (Wagaj and Patil, 2015)

In summary, the image segmentation, image enhancement and image representation are the three main process in digital image processing. With the respect to the work presented in this thesis the image segmentation, enhancement and representation were applied as the pre-processing process. The next section is 2.4 image classification is discussed.

2.4 Image Classification

The basic information of image classification is presented in this section. With the respect to this work in this thesis the image classification mechanism is commenced. The fundamental idea of data mining is presented in Sub-section 2.4.1. Sub-section 2.4.2 suggests the detail of image classification and finally, Sub-section 2.4.3 presents the brief detail of classification learning methods.

2.4.1 Data Mining

Data mining is the process of discovering and searching insightful, interesting, useful, as well as descriptive, understandable and predictive models from large-scale data (Zaki and Meira, 2014). Data mining considered as the non-trivial of novel, implicit and actionable knowledge from large datasets. The alternative terms refer to knowledge mining from database, knowledge extraction, data analysis and data archaeology.

In addition, data mining is considered as the core element in the knowledge discovery in database (KDD) process .The phase knowledge discovery in database was suggested in the first KDD workshop in 1989. The KDD processes including data selection, data cleaning, data transformation, pattern searching (data mining), patterns/finding evaluation, data visualization and knowledge representing.

In (Weiss and Davison, 2010) KDD consists of five processes: (i) Selection, (ii) Preprocessing, (iii) Transformation, (iv) Data mining and (v) Interpretation and evaluation as shown in Figure 2.15. From the Figure 2.15 can be observed that the KDD is a five-process mechanism. During selection process a large collection of data is selected. Once data selection is completed, the transformation process is then concerned in order to prepare appropriate data format to be ready for the data mining process (Weiss and Davison, 2010). During the data mining process, the data mining algorithms or techniques are implemented. When the pattern or hidden information is delivered the interpretation and evaluation process is applied to get the accurate knowledge.

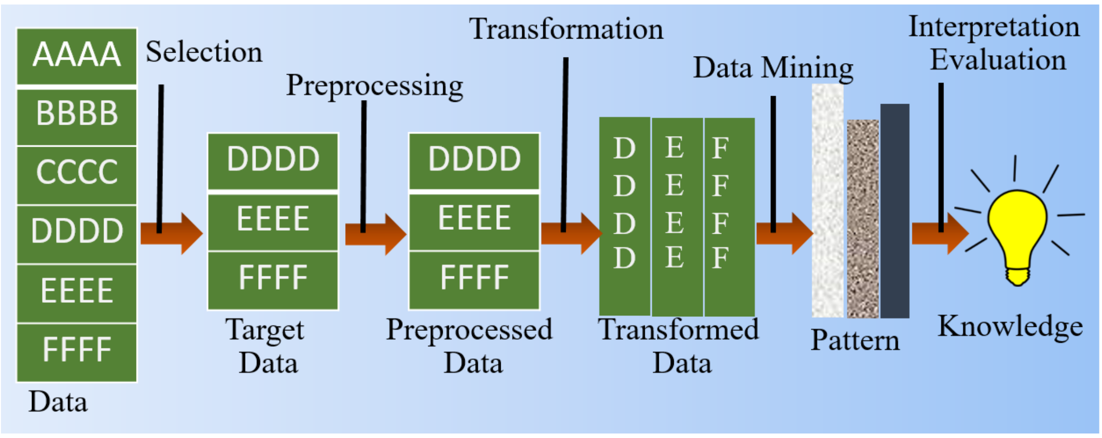


Figure 2.15 The example of KDD process for producing the knowledge

Data mining can be divided into two categories: (i) predictive data mining and (ii) descriptive data mining. Predictive data mining concerns with the prediction of behaviour based on historical data. In contrast descriptive data mining is directed at the discovery of patterns in existing data in order to use as the guideline for making the decision in the future.

In (Fu Yongjian, 2016) presented the tasks of data mining. In general data mining task is divided into three main categories: (i) Classification, (ii) Association and (iii) Clustering.

Classification is the process of finding a model that describes the data classes or concepts based on its properties. The aims of classification is to be able to use the derived model in the prediction of unknown objects. More spacifically, classification model can be used to classify the unknown class or future object and develop the classes of the objects in the database in order to get more understanding.

Association is the discovery which use to find the familiarity of identified patterns that are frequently occur together. Association rule reveals the associative relationship among the objects (Fu Yongjian, 2016) for example: Tesco Lotus Phuket generates an association rule that shows that 50% of time milk is sold with bread and only 20% of times biscuits are sold with bread.

Clustering is the process to group of similar kind of objects. Clustering analysis refers to forming group of objects that are very similar to each other. On the other hands, they are highly different from the objects in other clusters.

In the context of image data, image mining is then suggested. Image mining is one of a form of data mining which dealing with the extraction of implicit knowledge, image data relationship or other patterns not explicitly stored in a collection of images (Burl, et al., 1999). Image mining consists of multiple components (Burl, et al., 1999) such as: image analysis, image classification, image indexing, image retrieval and data management. The component of image mining is presented in Figure 2.16.

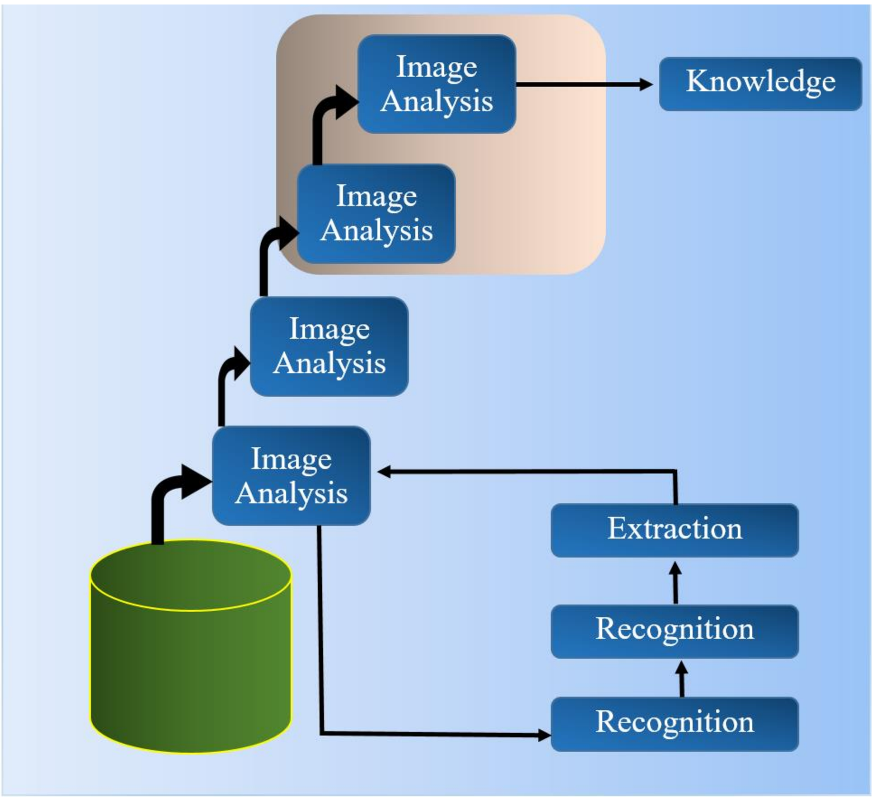


Figure 2.16 Traditional image mining process

2.4.2 Image Classification

Image classification is a branch of image analysis. It refers to classify an image by separate or group the similarity properties from the image into the same category.

With respect to digital image the image classification technique is applied to assign digital image to classes or categories using the image properties. Image classification is also widely used in medical image and remote sensing. Pattern recognitions is another cornerstone in computer science to help image classification classify an image, pattern recognition can be finding the similarities or patterns among small, decomposed problems that help to classify the complex image. In general, pattern recognition consists of three steps: (i) spectral pattern recognition (use spatial context to distinguish between different classes), (ii) spatial pattern recognition (used spectral reflectance in the different waveband to distinguish between different classes) and

(iii) temporal pattern recognition (use variations in image over time to distinguish between different classes).

Image classification process, the initial step is a collection of image (known-classes) is used as input data to generate the classification model which can be used to predict the class of unseen data in the future.

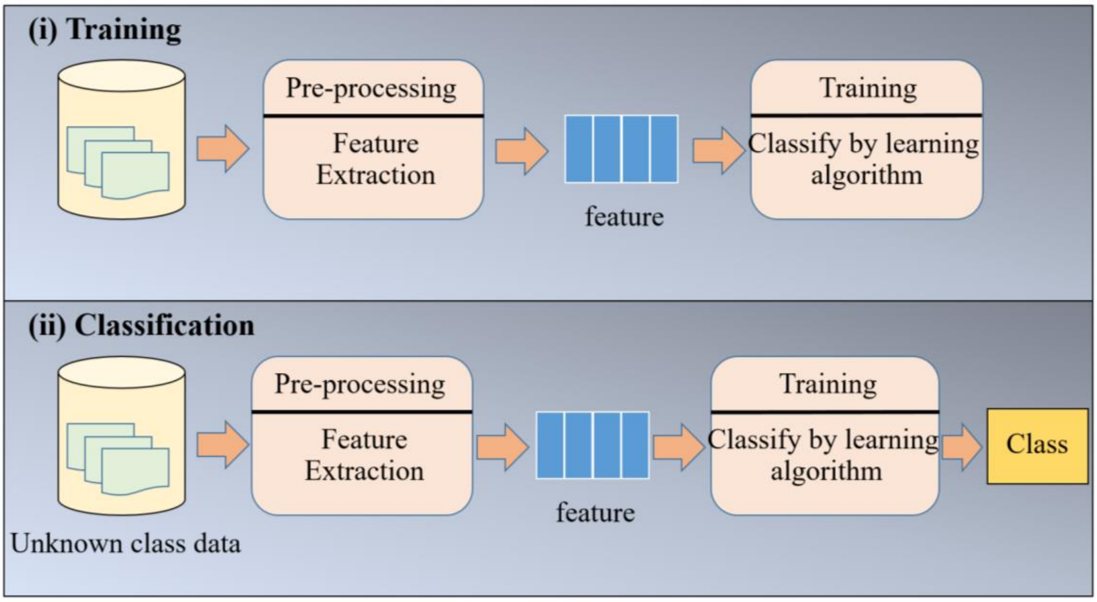


Figure 2.17 Image classification process

Image classification process is presented in Figure 2.17. From the Figure 2.17, image classification is a two sub-process: (i) training and (ii) classification.

During the training process a collection of image data is applied using pre-processing process to enhance the quality of image and segment the area of interest. The feature extraction is then applied resulting in a feature vector format. The next step is a classifier generation, which classification learning methods are applied to construct the desired classifier. The training data set used in training process, the pre-label record that typically represented by using an n-dimensional feature vector representation which in turn a set of attribute values and a class label $\{i_1, i_2, \ldots, i_{n-1}, c_n\}$ where i_i is an attribute value and c_n is a class label that $c_n \in C$.

Once the classifier has been generated, the second process is classification. In this process the generated classifier from training process is applied to the unknown/unseen image data in order to classify/predict the data class.

As noted above image classification technique consist two main processes where in the classification application process need to predict on the usage of discrete class label.

2.4.3 Classification Learning Methods

As mentioned before the work presented in this thesis is concerned to image classification. There are nine classification learning methods are presented in this sub-section: (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.

Decision Tree

Decision Tree (Quinlan, 1986) is considered as the most popular and widely used for classification learning methods. A decision tree is a mechanism to classify which is a recursive partition a collection of instances. Root tree is considered as the top node of the decision tree. In other words, decision tree is a classifier which instance by sorting them downs the tree from the root to some leaf node (also known as terminal or decision nodes). Root node is a direct tree with a node that has no coming edges. Every other node has exactly one coming edges. An internal or test node is the node that outgoing edge. In general, test node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. Figure 2.18 illustrates the decision tree for the direct mailing response.

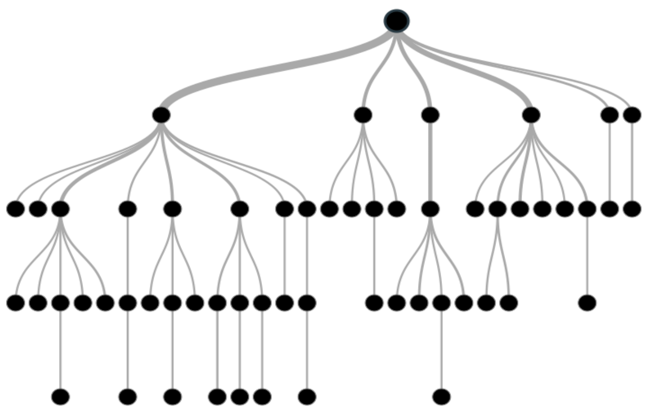


Figure 2.18 Response to Direct Mailing Decision Tree

Binary Split Tree

Binary split is the improvement of decision tree. Each node of binary split tree contains only two values (binary 0-1). The purpose of designing binary split tree is to store the statistic datasets. The split tree is a subset of decision tree while each node of a decision tree contain multiple values. The further search in the tree is used by the split value in the split while the key value is not matched to the search value. The example of binary split tree is presented in Figure 2.19:

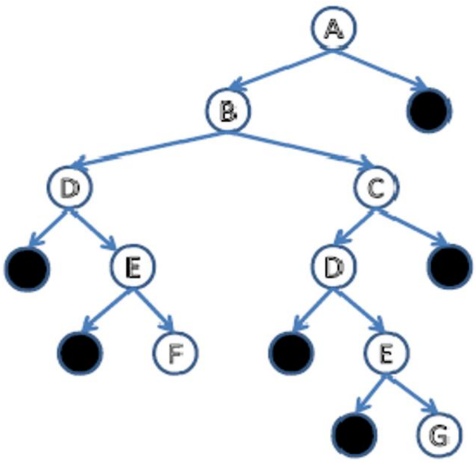


Figure 2.19 Binary Split Tree

Average One-Dependence Estimators

Average One-Dependence Estimators (AODE) is presented in this subsection. AODE is a probabilistic classification learning technique, AODE has improved from the Naïve Bayesian classifier (Webb, et al., 2005) by addressing the problem of attribute-independence. For example, in the class of y, which has a set of features $x_1,..., x_n$, AODE can be applied to find the probability of each class y by using the following equation:

$$\hat{P}(y|x_1, \dots, x_n) = \frac{\sum_{i:1\leq i\leq n \wedge F(x_i)\geq m} \hat{P}(y,x_i) \prod_{j=1}^{n} \hat{P}(x_i|y,x_i)}{\sum_{y'\in Y} \sum_{i:1\leq i\leq n \wedge F(x_i)\geq m} \hat{P}(y',x_i) \prod_{j=1}^{n} \hat{P}(x_i|y',x_i)} \quad 2.3$$

where

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

F is the frequency

m is a user specified minimum frequency.

Bayesian Network

Bayesian Network (BN) or Probabilistic Networks (PNs) is discussed in this subsection. BN is consider as a graphical probability model which can be used for reasoning and the decision making in uncertainty (Friedman, et al., 1998). In addition, the BN have been applied as a directed acyclic graph (DAG) and each node of BN represents a domain variable or dataset attribute. BN formally work depend on the Bayes rule while the The Bayes' rule can be written as the equation below:

Assume, there are Ai attributes where i= 1,...,n, and take value a_i where i= 1,...,n . Assume there is C as class label attribute and U= (a_1,..., a_n) as unclassified test instance. Hence, U will be classified into class C based on Bayes rule is represented as:

$$P(C|U) = \arg\max P(U|C) \qquad 2.4$$

In addition, In work (Watt and Bui, 2008) presented the Bayesian belief Network (BBN) model which is literally the same to Bayesian Network, applied to knee OA detection. The graph model of BBN for knee OA detection is shown in Figure 2.20:

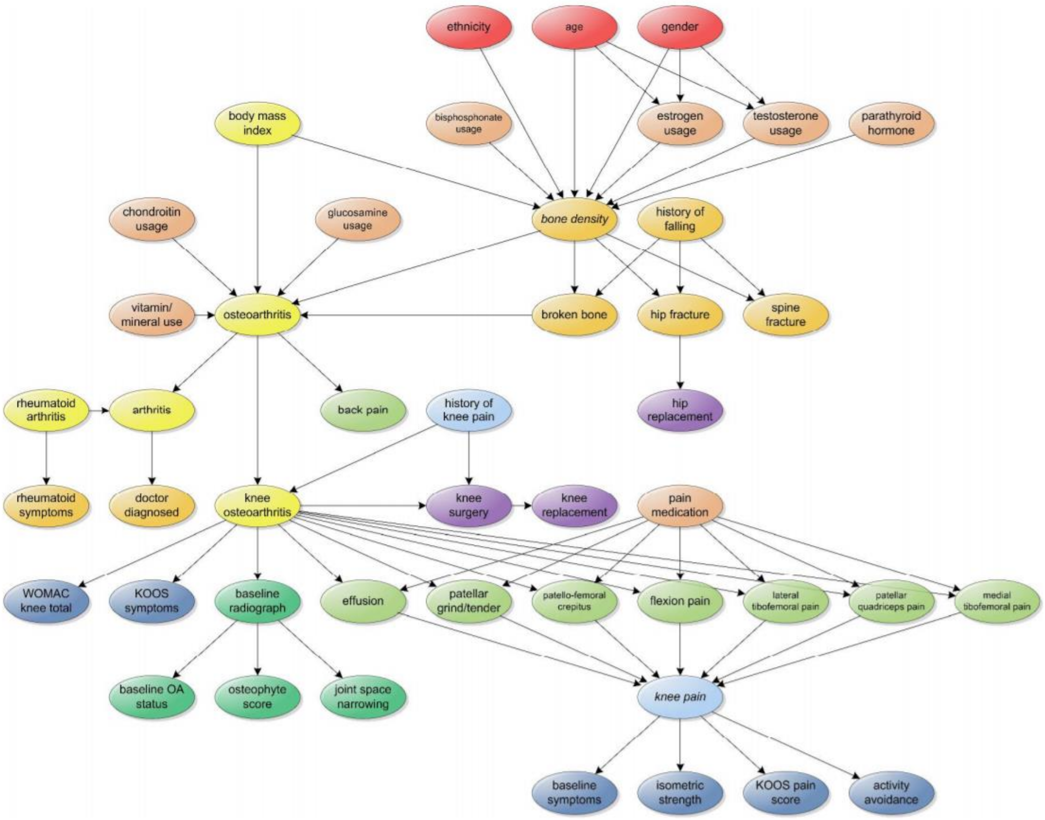


Figure 2.20 Bayesian belief network for knee osteoarthritis (Watt and Bui, 2008)

Naïve Bayes

Naïve Bayes has been widely implemented in clustering and classification. Naïve Bayes sometime called idiot Bayes, simple Bayes or Bayes classifier. The Bayes theorem is used in Naïve Bayes for prediction both in classification and clustering. The Bayes theorem is shown as the equation 2.5 below:

$$P = \frac{p(d|\ c_i)\ p(c_i)}{P(d)} \qquad 2.5$$

where

$p(c_i|d\ )$ is the probability of instance d being in class $c_{i,}$

$p(d|\ c_i)$ is probability of generating instance d given class $c_{i,}$

$p(\ c_i)$ is the probability of occurrence of class c_i

p(d) is probability of instance d occurring.

Support Vector Machine

Support Vector Machine (SVM) is a popular classification method that has been widely applied for the classification task. The main used of SVM is to separate instances of two classes by constructing an N-dimensional hyperplane amount of two training sample classes in the feature set for the most optimal way (Cortes and Vapnik, 1995). SVM classifiers formally divided into two groups: (i) linear and (ii) non-linear.

- Linear Classification

  In 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 ∈-1,+1, i = 1,...,n, can be written as the equation below:

$$\begin{cases} x_i.\omega + b\ \geq +1 & for\ y_i = +1 \\ x_i.\omega + b\ \leq -1 & for\ y_i = -1 \end{cases} \qquad 2.6$$

For non-linearly separable, the SVM is defined as below equation:

$$\begin{cases} x_i.\omega + b \ \geq +1 - \ \xi_i \quad for\ y_i = +1 \\ x_i.\omega + b \ \leq -1 + \xi_i \quad for\ y_i = -1 \\ \xi_i \geq 0, \quad i = 1, n \end{cases} \qquad 2.7$$

- Non-Linear Classification

In non-linear SVM classification, SVM is defined as:

$$f(x) = \sum_{i=1}^{n_s} \alpha_i y_i \, P(x_i, x) + b \qquad 2.8$$

where

n_s is the number of support vector.

α is non-negative Lagrange multipliers

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

In addition, with respect to the application of knee OA classification was applied LibSVM in Waikato Environment for Knowledge Analysis (WEKA) with Radial basis function kernel (Gaussian kernel) that is the default kernel in WEKA and C-SVC operator default value 0. The default value of equation of RBF kernel on two sample x and y, represented as feature vectors in sample input, is defined as:

$$K(x, y) = \exp[-\gamma(x - y)^2] \qquad 2.9$$

where

$(x\text{-}y)^2$ is considered as the squared Euclidean distance between the two feature vectors

$$\gamma = \frac{1}{2\sigma^2} \qquad 2.10$$

where

σ is a free parameter

Logistic Regression

Logistic regression is a statistical method which used to analyse a dataset. The dataset can be one or more independent variables which determine an outcome. The outcome is measured with a dichotomous variable (dependent variable normally is binary or dichotomous). 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 2.11, where p is the probability of presence of the characteristic of interest.

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \ldots + \ldots b_kX_k \qquad 2.11$$

Sequential Minimum Optimization

Sequential Minimal optimization (SMO) is a computer algorithm which is used to solve the quadratic programming (QP) problem that happen during the training of support vector machines (SVM). In order to solve 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. In other words, SMO algorithm can search with feasible region of two kind or type of problem then maximize the objective function. In other words, by using Lagrangian, this QP problem can be converted into a dual where the objective function Ψ is solely dependent on a set of Lagrange multiplier α_i. The equation of Lagrange is presented in Equation 2.12:

$$\min \Psi\,(\vec{\alpha}) \equiv \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j \qquad 2.12$$

where

N is the number of training examples

The SMO use Radial basis function kernel (RBF kernel) for classify the task.

Back Propagation Neural Network

Back Propagation Neural Network or Back Propagation is a computer algorithm/system which sets the model design as the human brain and nervous system. Back propagation is a mechanism used in artificial neural networks to calculate a gradient that is needed in the calculation of the weights to be used in the network. Back propagation is commonly used to train deep neural networks, (Nielsen and Michael, 2015) a term referring to neural networks with more than one hidden layer. Figure 2.21 is illustrated the general mechanism of Neural Network that has Back propagation to train in deep neural network.

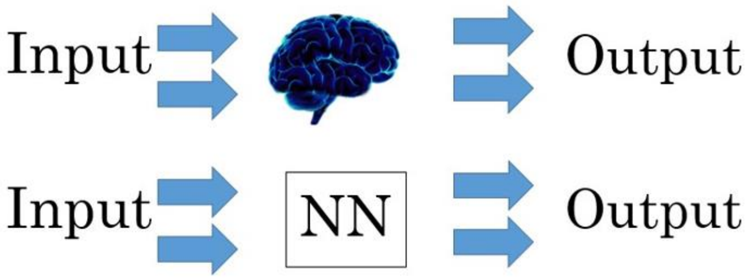


Figure 2.21 The example of Neural Network mechanism

In the Neural Network (NN) consists of three layers: (i) input, (ii) hidden, and (ii) output. For Back propagation is use to train in neural for learning. In other word, in the case of learning in NN, back propagation 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 sometime 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 back propagation neural network known as the Multilayer Perceptrons 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 2.22:

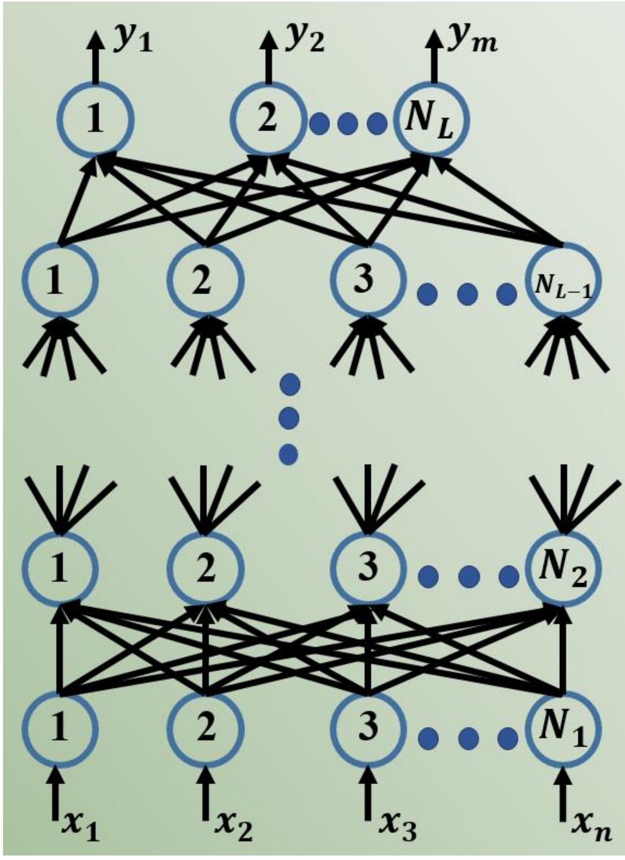


Figure 2.22 The example of Back Propagation Neural Network

2.5 Feature Selection Methods

In this section feature selection is described. Feature selection consider as the main process in the research study in order to make a feature vector which is very important for the classification process. The main use of feature in the research study is to reduce data dimensionality for making the feature vector. In the study, there are five feature selection techniques are applied, including: (i) Correlation-based Feature Selection (CFS), (ii) Chi-Squared, (iii) information gain, (iv) Gain Ration, and (v) Relief feature selection.

Correlation-based Feature Selection

Correlation-based feature selection is a techniques used to reduce the feature space that applied a heuristic search for evaluation the worth of feature subsets. In work (Hall, 1999 and Lewandowski, 2015) reported that CFS used the heuristic search for calculation the evaluation of feature subsets that based on the hypothesis "Good feature subsets contain features highly correlated with the classification, yet uncorrelated to each other". In order to reduce the feature space, CFS used Symmetric Uncertainty. The fooling equation is defined as the Symmetric Uncertainty equation of two nominal attributes A and B:

$$U(A,B) = 2\,\frac{H(A)+H(B)-H(A,B)}{H(A)+H(B)} \qquad 2.13$$

where

H is the function of entropy

With the reference to Equation 2.13, CFS can be forms as the equation below:

$$CFS = \frac{\sum_{i=1}^{n} U(A_i,C)}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} U(A_i,A_j)}} \qquad 2.14$$

where

C is the class of the feature

(A_i,A_j) is the indicates a pair of attributes

Chi-squared

Chi-squared is used to measure the relationship of dependency between a feature and a class (Plackett and Pearson, 1983). Chi-squared is represented as χ^2 symbol and defined as the equation below:

$$\chi^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad 2.15$$

where

O_{ij} represent as the observed frequency

E_{ij} represent as the expected frequency

Information Gain

Information gain or IG is considered as one of the popular feature selection techniques. The implementation of IG is for selecting the test attribute at each node. IG can measure the number of information in bits to the class prediction while the available information present as a feature and class distribution (Lei, 2012). For instant, if IG of a feature *s* directly related to a collection aspects B, thus IG can be represented as the equation below:

$$IG(B, s) \sum_{v \in Values(s)} \frac{|B_v|}{|B_s|} Entropy(B_v) \qquad 2.16$$

where

Values(s) presents as the set of all possible feature values *s*

$|B_v|$ presents as the cardinality to the subset of class related to feature *s*

$|B_t|$ presents as the cardinality to the set of aspects belonging to feature *s*

Gain Ratio

Gain ration (GR) is presented in this sub-section, an extension of IG is considered as gain ration technique. The implementation of decision tree in IG for one reason is to select the test attribute of each node (Han, et al., 2012). Thus, the applying of GR to make IG better performance by choosing an attribute by token number and size of branches. In other words, GR is implemented to IG in order to reduce the bias of IG on each branch. GR is written as the equation below:

$$GR(attribute) = \frac{Gain(attribute)}{Entropy(attribute)} \qquad 2.17$$

Relief

Relief feature is considered as the last technique is presented in this sub-section. Relief first introduce in work (Kira and Rendell, 1992). Relief is the weight based algorithm that the related features are considered as the one who has a better distinction between the classes (Kira and Rendell, 1994). In order to understand the relief, a sample of dataset is selected, the nearest neighbouring sample that belongs to the same class is called 'Near-hit'. On the contrary, the nearest neighbouring sample that belongs to the opposite class is called 'Near-miss'. The relief measure these two weights: (i) Near-hit and (ii) Near-miss. In addition, all reiterations of M times, relief takes the feature vector of X that belong to random sample of Near-hit and Near-miss. After M iteration has gone, relief separates each item of the weight vector by M, thus, the relevance vector is created. Finally, the selected feature gets from the features that have relevance value greater than a threshold *T* value. The Relieff of Near-hit and Near-miss feature are illustrated in Figure 2.23 below:

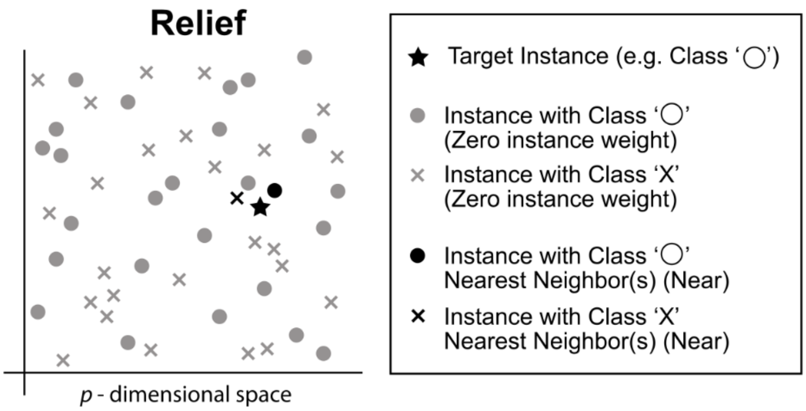


Figure 2.23 The Example of ReliefF Near-miss and Near-hit (Ryan, et al., 2017)

2.6 Deep Learning using Convolutional Neural Network

The implementation of deep learning in medical image analysis is presented in this sub-section. Deep learning is one of the mechanism in machine learning which is directed to the algorithm of brain structure and function called artificial neural networks. In other words, deep learning is considered as a technique developed from the artificial neural network which was inspired from human brain neurons connected system. The application of deep learning has been widely used in computer vision technology including: image classification, object detection, and image segmentation. Deep learning model is known as the well-known applications of deep learning, of those deep learning model is Convolutional Neural Network (CNN) (LeC, et al., 1999). CNN is a deep learning techniques that used for image classification. With the respect to the purpose of study, CNN have been applied to various type of medical image. For example, the latest study have been shown that deep learning are very efficient to medical image approach [Esteva, et al., 2017; Anthimopoulos, et al., 2016; Cole, et al., 2017; Havaei, et al., 2017 and Kooi, et al., 2017). The Implementation of CCN to brain MR image for segmented brain into a number of classes is presented in (Moeskops, et al., 2016). In work (Pereira, et al., 2016) presented the brain tumours segmentation with MR image with small 3x3 kernel of CNN. For knee application, in work (Prasoon, et al., 2013) have presented the knee tibia cartilage segmentation from MR image scan and tested on 114 unseen scans. The further information of CNN used with respect to work in this thesis is discussed in Chapter 6.

2.7 Evaluation and Measurement

To evaluate the classification performance for the generated classifier, the evaluation measure is applied. With the respect to the work presented in this thesis, the classification performance was recorded in term of (i) Accuracy (AC), (ii) Sensitivity (SN), (iii) Specificity (SP), (iv) Precision (PR), and (v) F-Measure (FM). The evaluation measurement is discussed further detail as follow. The confusion matrix and some evaluation measures calculating from confusion matrix is presented in 2.7.1. The basic idea of AUC is discussed in 2.7.2.

2.7.1 Confusion Matrix

The confusion matrix is used to measure how well of applied data mining algorithm perform on a given data set. It can help to find out which data mining algorithm give the best performance or the worst performance. The confusion matrix is performed about the predict class (predict value) and actual class (actual value), Table 2.2 presents the confusion matrix.

Table 2.2 Confusion Matrix

		Predicted Class	
		True	False
Actual Class	True	TP	TN
	False	FP	FN

From the Table 2.2, TP stands for True Positive rate used to measure the proportion of correctly identified positive test records. TN refers to the True Negative which used to measure the proportion of correctly identified negative test records. The FP is False Positive rate which used to measure the proportion of incorrectly identified positive test records and FN stands for False Negative which use to measure the proportion of incorrectly identified negative test records. There are a number of evaluation measures based on the confusion matrix including: (i) Accuracy, (ii) Sensitivity, (iii) Specificity, (iv) Precision, and (v) F-measure. The equation for each measure is presented in the Equation 2.18 to Equation 2.22, respectively.

$$Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \qquad 2.18$$

$$Sensitivity = \frac{TP}{(TP+FN)*100} \qquad 2.19$$

$$Specificity = \frac{TN}{(TN+FP)*100} \qquad 2.20$$

$$Precision\ = \frac{TP}{TP+FP} \qquad 2.21$$

$$F - measure = 2\,\frac{Precision*Recall}{Precision+Recall} \qquad 2.22$$

where

$$Recall = \frac{TP}{TP+TN} \qquad 2.23$$

2.7.2 Area Under Curve

Area under curve (AUC) is considered as the best method for indicating the overall quality of the classifier (Fawcett, 2006). The AUC can be calculated from Receiver operating characteristic (ROC) graph. ROC graph is a two-dimensional plot of the TP rate as the y-axis and the FP rate as the x axis (Fawcett, 2006). The ROC graph is shown in Figure 2.24.

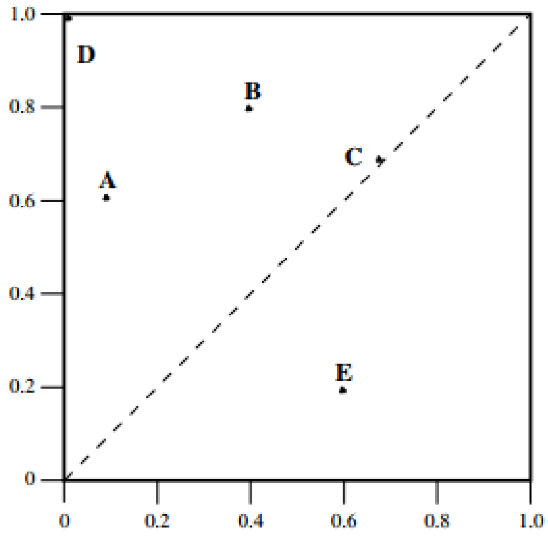


Figure 2.24 The ROC graph with five discrete classifiers (Fawcett, 2006)

From Figure 2.24 the lower left point (0, 0) is the point that a classifier perform no false positive errors but also gets no true positives. On the contrary, the upper right point (1, 1) is the point that a classifier perform full false positive errors but also gets full true positives. Thus, amounts five classifiers A-E, D’s perform the best one due to the full of true positive rate.

AUC refers to the area under the ROC curve. Figure 2.25 presents the AUC in ROC graph which the random value from 0.0 to 1.0 in probabilistic classifier.

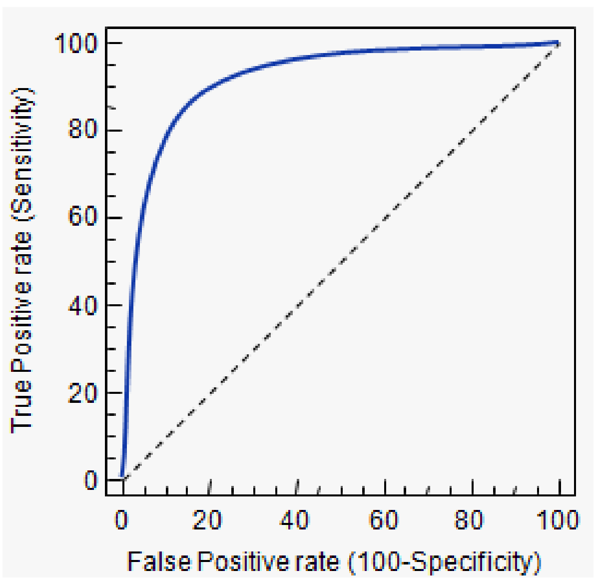


Figure 2.25 Area Under ROC Curve (AUC) (Fawcett, 2006)

2.8 The Comparison of related work

This section the comparison of previous or related works, typically OA detection, is discussed in this section. Table 2.3 presents the comparison of related works in this research study:

Table 2.3 The comparison of related works

Previous work	**Image Types**	**Osteoarthritis detection**	**Osteoarthritis Stage Detection**	**Accuracy/result, methodology**
Kubakaddi, et al., 2013	MRI		Just classified (KOA, Normal and Doubtful KOA)	Calculate the thickness of cartilage (Shape analysis).
Pandey, et al., 2016	X-ray	Detect by shape analysis		Calculate the thickness of cartilage with the rate of result is higher than 65.00% (Shape analysis).
Gornale, et al., 2015	X-ray	OA/ Non OA detection by using different technique of classification		Result provides 87.92 % by using different features calculations such as shape, statistical, first-four moment texture and Haralick texture.
Jin, et al., 2012	Infrared	OA/Non OA by extraction the patella-centering of knee feature.		Result provides 85.49% of accuracy rate, 85.72% of sensitivity and 85.51% of specificity buy using SVM classifier.

2.9 Summary

In conclusion, many kinds of medical images have been applied to diagnose OA disease with image processing technique as mentioned above. As an example of work (Pandey, et al., 2016; Shamir, et al., 2009a, 2009b, 2010; Gornale, et al., 2015; Shamir, 2011; Herman, et al., 2015; Kellgren, et al., 1963; Kellgren and Lawrence, 1957 and Kubakaddi, et al., 2013) have been applied the to various of medical image to analyse OA/non-OA and OA stage using image classification and image processing technique. To summarise of this chapter is to tell a brief of each previous work had implemented in image processing technique such as image enhancement, image segmentation, feature extraction, feature selection, evaluation and measurement. Some techniques for image processing. There were some brief study for machine learning algorithms for example, neural network, decision tree and logistic regression could be used for classification process. The detailed of each process work in medical X-ray image analysis is discussed in Chapter 3.

CHAPTER 3

KNEE X-RAY IMAGE DATASET

3.1 Introduction

With the respect to the study of this research, the knee X-ray image data sets were used. The data sets were applied to two different studies: (i) knee OA detection where the collected image is divided in two groups (normal control and OA case) and (ii) knee-OA stage classification where the collected image is divided into five groups (*stage* 0 till *stage* 4). The rest of this chapter is organised as follow: the discussion of knee X-ray image collection used in this thesis is presents in Section 3.2. The labelling process for image data sets is discussed in Section 3.3.The description of region of interest (ROI) and image enhancement are commenced in section 3.4. Finally the whole chapter is concluded in section 3.5.

3.2 Knee X-ray Image Collection

A collection of knee X-ray image data set used in this work is discussed in this thesis. A total 131 images were obtained from two local hospitals: (i) Bangkok Hospital (Phuket branch) and (ii) Dibuk Hospital. The image used with respect to the work presented in this thesis was taken in Anterior Postero (AP) position from random age, both male and female, and both left and right. The image was obtained in greyscale colour as Digital Image and Communication on Medicine (DICOM) format. Figure 3.1 shows the AP position original image with DICOM format which the image size was 1416 x 138 pixels.

Figure 3.1 Original X-ray Image in DICOM format which view in DICOM Viewer Tool

The obtained image sometime were collected in different positions. The examples X-ray image in Figure 3.2 were eliminated from our data set.

Figure 3.2 Unselected image

Form the Figure 3.2 it can be seen that, the images were eliminated from the data set presented in Figure 3.2 (a) the image toke in different PA position could not be considered for the study. In Figure 3.2 (b) there is some steel connected with the bone. Thus, in this case the knee-OA cannot analyse because this is not the nature of bone. Finally, Figure 3.2 (c) was eliminated due to some technical error while producing the knee screening image like unclear image.

To prevent the privacy, the collected image were focused only on the image, no personal information was given from the hospital.

3.3 Label Image Data Set

As noted in Chapter 1 Section 1.3 research methodology, there were two different studies had been carried out with respect to the work in this thesis: (i) knee-OA detection and (ii) knee-OA stage classification. Therefore the image data sets were labelled in different ways.

With respect to knee-OA detection, the image data set was separated into two different classes by expert domain from Bangkok Hospital. The classes were: (i) image with OA and (ii) normal control (image without OA) that were presented in Figure 1.1 of Chapter 1. Recall from the Figure 1.1 it illustrated that Figure 1.1 (a) presented the knee without OA, in this figure we can see image with the clear joint space, while Figure 1.1 (b) illustrates the knee with OA, in this figure the knee joint space was disappear from the connected bone. In the study of knee-OA detection, a collected image of 131 images were divided into (i) 63 of normal control image and (ii) 68 of OA image.

With respect to knee-OA stage classification, the Kellgren and Lawrence (Kellgren, et al., 1963 and Kellgren and Lawrence, 1957) system have level the stage of knee-OA into five different stage (*stage* 0 till *stage* 4) as presented in Figure 1.2 of Chapter 1. Recall from Figure 1.2 it can be seen that Figure 1.2 (a) presented the *stage 0* or normal control, in this stage there no sign of OA has appeared. Figure 1.2 (b) pictures about the *stage* 1 of knee-OA or doubtful stage, in this stage a slightly sing of OA has appeared such as doubtful joint space narrowing. ; Figure 1.2 (c) presented the *stage* 2 or Mild grade of knee-OA stage, in this stage the possible joint narrowing was occurred. Figure 1.2 (d) illustrated the *stage* 3 or the moderate grade of knee-OA stage, in this stage the joint space reduction and deformity of bone contour were appeared.

Lastly, the Figure 1.2 (e) presented the *stage* 4 or severe stage, in this stage the grate pain when movement was occur to patient and thesis no any joint space more. For knee-OA stage classification study, the 131 collected image of were classified by expert domain from Thongsoung Hospital (Dr. Chaowakon Saehang, MD).

There are three main ways to be consider of knee-OA stage classification based on Dr. Chaowakon with his professional work more than 10 years. These ways include: (i) joint space, (ii) Lateral or Medial condyle, and (iii) the space of Tibia and Femur around the joint space. In Figure 3.3 is illustrated the three ways of OA stage classification by Dr. Chaowakon.

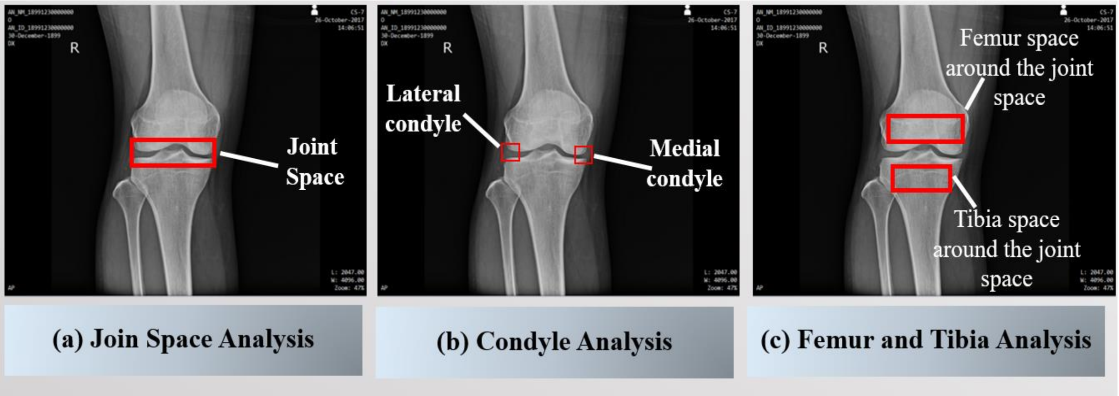


Figure 3.3 Three Ways of OA stage Classification

From Figure 3.3 there were three sub-images are illustrated. The Figure 3.3 (a) illustrates the joint space between Femur and Tibia bone (the large and clear join space consider as normal control image but if there is no any joint space it considers as the *stage* 4 of knee-OA). Figure 3.3 (b) illustrated the way to define knee-OA stage by focusing on both condyle (medial and tibia condyle). If the condyle look not clear and look like brittle, it would be consider to have OA. Finally, Figure 3.3 (c) illustrates the tibia space and femur space around the joint space based on the bone texture analysis.

With respect to the ways that presented by Dr. Chaowakon illustrated in Figure 3.3. Dr. Chaowakon had divided the 131 X-ray images data set in to five group of OA stages as the table below:

Table 3.1 The image number of each OA stage

OA Stage	*Stage* 0	*Stage* 1	*Stage* 2	*Stage* 3	*Stage* 4
No. of Images	39	19	21	44	8

**Note that: The data set consist of 131 images, due to shape of the bone is not clear I had removed three images of: (i) one image from OA *stage* 0 and (ii) two images of *stage* 1. Thus, this set of images is further use shape based presented in Chapter 5 and Chapter 6 respectively.

3.4 Region of Interest (ROI) Segmentation and Enhancement

The mechanism of region of Region of Interest identification and image enhancement are presented in this section. The objective of this process is to identify ROIs from the given image data set and improve the quality of the image so that the feature extraction could be commenced. Sub-section 3.4.1 presents the detail of ROIs used in this thesis. The mechanism used for ROIs quality improvement is discussed in Sub-section 3.4.2.

3.4.1 Region of Interest

With respect to the work presented in this thesis, the mechanism of identifying ROIs is presented in this sub-section. ROIs identification was done manually. There are two mechanisms for ROIs identification:

SET A: Four ROIs were identified for texture analysis (Chapter 4). The four ROIs were used in texture analysis include: (i) lateral femur, (ii) medial femur, (ii) lateral tibia, and (iv) medial tibia. Figure 3.4 illustrated the identification of four ROIs.

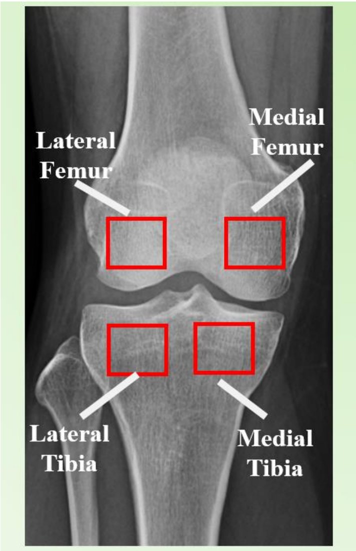


Figure 3.4 Four different ROIs

SET B: Joint space region was identified for graph-based (Chapter 5) and Convolutional Neural Network (Chapter 6). With respect to graph-based and CNN study purposed, the ROIs identification was focused on knee joint space. There are two ROI were identified: (i) the whole knee ROI, in this ROIs segmented the whole knee cover both tibia and femur region and (ii) knee joint space ROI, in this ROI segmented only the region around joint space. The whole knee ROI and knee joint space ROI are presented in Figure 3.5 below:

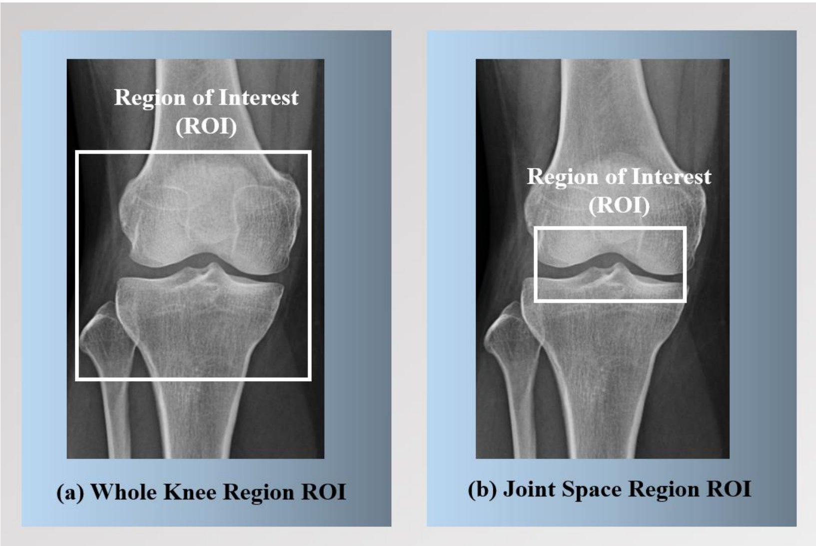


Figure 3.5 ROIs of joint space area

With reference to Figure 3.5, there two ROIs are presented in shape based approach. Figure 3.5 (a) illustrated the whole knee ROI, in this region cover of condyle and region around the joint space. Figure 3.5 (b) presented the joint space. The segmentation of ROI were presented in this section. Next, the enhancement techniques is applied in joint space ROI.

3.4.2 ROI Image Enhancement

With respect to the work presented in this thesis, the mechanism of identifying ROIs is presented in this sub-section. ROIs identification was done manually. There are two mechanisms for ROIs identification:

Once the ROIs had been identified, the image enhancement could be performed. With respect to the work presented here, the Otsu thresholding was applied to the obtained ROIs. The detail of Otsu was presented in Chapter 2 Section 2.3. The objective of the process is to improve the quality of image, thus to be ready in the feature extraction process. The Otsu thresholding was applied only in ROIs of joint space area to make the joint gap clearer in order to facilitate in shape analysis. The Figure 3.6 is represented the example of Otsu operator:

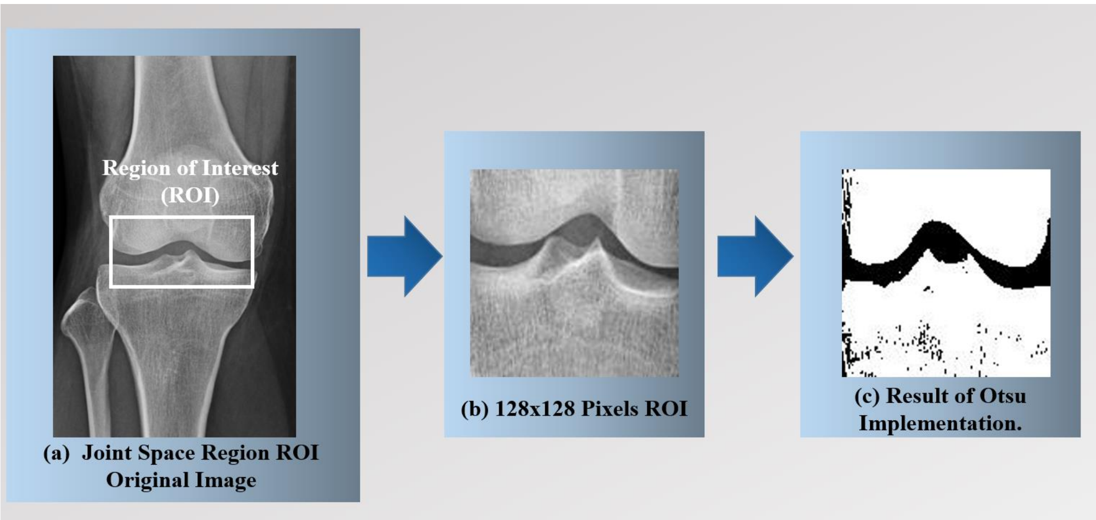


Figure 3.6 The example of Otsu operator to knee joint gap area

From Figure 3.6 it can be seen that the original image with knee joint gap ROI identification is presented in Figure 3.6 (a). Figure 3.6 (b) illustrated the

128x128 pixel ROI of knee joint gap area and Figure 3.6 (c) illustrated the result Otsu operator applying to Figure 2.3 (b).

3.5 Summary

In summary, this chapter presented on X-ray image data set that used in this thesis work. The chapter focused the data set collection and the ROIs identification for each image of data set that is suitable to use in the next process of the research work. The way to label the data set for use in the research was presented in this chapter. Finally, pointed out the enhancement technique for improvement the ROI quality.

CHAPTER 4

OSTEOARTHRITIS CLASSIFICATION USING KNEE X-RAY IMAGERY: TEXTURE-BASED APPROACHR

4.1 Introduction

The introduction of the texture-based approach for knee-OA classification is discussed in this section. While the graph-based approach using quadtree decomposition and the Convolution Neural Network are presented in Chapter 5 and Chapter 6 respectively. With respect to research methodology presented in Chapter 1 Section 1.3, there were two studies: (i) knee OA detection and (ii) knee-OA stage classification.

For both study (knee-OA and knee-OA stage), the implementation of texture based approach for OA classification is considered. The main purpose of knee-OA detection study is to classify OA and non-OA (normal), while knee-OA stage classification is to classify the stage of OA. In the context of texture analysis, image data SET A (As mentioned in Chapter 3 Sub-section 3.4.1 Region of Interests) was used. A total of 131 images was used for knee-OA detection study the dataset comprise of two groups: (i) 68 images of OA case and (ii) 63 images for normal control (the data set was divided by expert domain from Bangkok Hospital). SET A data set was divided in five groups in the context of OA stage classification : (i) 39 images of *stage* 0, (ii) 19 images of *stage* 1, (iii) 21 images of *stage* 2, (iv) 44 images of *stage* 3 and (v) 8 images of *stage* 4.

To be more specific, for both studies in this section ten image texture feature descriptors were applied to extract the potential feature from each ROI. ROIs were identified, include: (i) Later Femur (LF), (ii) Medial Femur (MF), (iii) Lateral Tibia (LT), and (iv) Medial Tibia (MT). The detail of ROIs identification was presented in Chapter 3 Sub-section 3.4.1. An isolate ROI was applied to ten different texture descriptors that are presented in detail in Section 4.2 consider as the next section. As noted in Section 2.5 Feature Selection Methods of Chapter 2.The implementation of texture descriptors to four ROIs or sub-images produced a groups of feature space. In amount of feature spaces contain a lot of features, in these features comprise of different feature value and properties, so the selection of feature is very important to produce the good feature vector which this technique was called Feature Selection Technique (As mentioned in Chapter 2). In similar vein, feature selection is a major technique to reduce feature dimensionality in feature spaces. In the knee-OA detection study, five dissimilar feature selection techniques were presented in Chapter 2 were applied. While in knee-OA stage classification applied only one feature selection technique that considered as the best performance of feature selection method from the knee-OA detection study. Make perfectly clear, a schematic of texture base approach is illustrated in Figure 4.1 below:

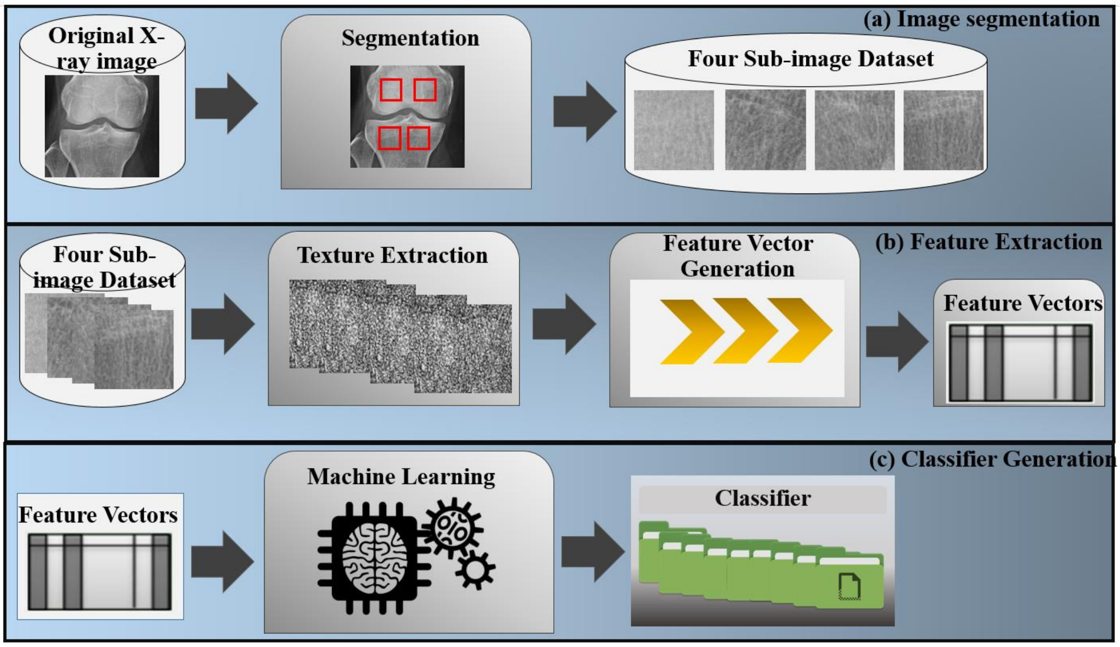


Figure 4.1 The Texture-Based of OA classification framework

From Figure 4.1, it can be seen that the texture based approach for OA classification comprises of three main processes: (i) image segmentation, (ii) Feature Extraction, and (iii) Classification Generation. The ROI identification was discussed in

Chapter 3 and will thus not discussed further in this chapter. Once the ROIs were obtained, the feature extraction is then command to transform the selected pixel of region into the appropriate from suitable that better for classifier generation. The classifier generation is the process of desire classification model construction using a generated feature (The second process). With respect to the work in this research, nine learning algorithms were applied.

The challenge of this work is how to extract the best saline information of the image based on texture analysis. With the respect to the work here, ten texture descriptors was applied to extract the information of ROIs (Texture descriptor presented in Section 4.2 below). One the ROI was extracted, the feature space was created. In order to make a good feature vector that is suitable to use forward with the learning classifier, the feature space reduction is then applied. In order to reduce the feature space in term of dimensionality and number the feature selection is applied. Thus, the performance of feature extraction process (The second process) that combine of texture descriptor and feature selection techniques to selected the feature vector consist of three sub-processes: (i) Texture Extraction, (ii) Feature Vector Generation, and (iii) Feature Vector Selection. Finally, the feature vector representation was used with learning classifier for the classification process (the third process).

The rest of this chapter is organised as follow: The information of texture descriptors techniques that used in the OA classification works are presented in Section 4.2. Section 4.3 reported the feature selection techniques and classification using in the proposed work. The evaluation of OA classification works are presented in Section 4.4, for knee-OA detection study was evaluated in Sub-section 4.4.1 and knee-OA stage classification was evaluated in Sub-section 4.4.2. The discussion of OA classification works are illustrated in Section 4.5. Finally, the summary of OA classification studies are presented in Section 4.6.

4.2 Texture analysis

In this section, the texture descriptors which were applied in the proposed framework is presented. There are ten texture descriptors were applied include: (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, further information of each texture descriptor is presented as follows:

4.2.1 Histogram feature

The grey image contains of histogram feature which is received by stage of the art of histogram based feature. Three are six features were applied in the study include: (i) Mean, (ii) Variance, (iii) Skewness, (iv) Kurtosis, (v) Energy and (vi) Entropy.

1) Mean is the average of feature j in the grey level of image. Mean can be defined as:

$$\mu = \sum_{j=1}^{n} jP(j) \qquad 4.1$$

where

P(j) is the probability of j, P(j) can be defined as the equation below:

$$P(j) = \frac{H(j)}{M} \qquad 4.2$$

M is the block number and H(j) is the histogram function

2) Variance is a measure of the histogram width that measures the deviation of grey levels from the Mean:

$$\sigma^2 = \sum_{j=1}^{n} (j - \mu)^2 \, P(j) \qquad 4.3$$

where

$\sigma^2$is the variance

μ is the mean

3) Skewness is a measure of the degree of histogram asymmetry around the Mean:

$$skew = \frac{1}{\sigma^3} \sum_{j=1}^{n} (j - \mu)^3 \, P(j) \qquad 4.4$$

where

σ is the standard deviation which is the square root of the variance presented in equation 4.3.

4) Kurtosis is a measure of the histogram sharpness.

$$Kurtosis = \frac{1}{\sigma^4} \sum_{j=1}^{n} (j - \mu)^4 \, P(j) \qquad 4.5$$

5) Energy is used to describe a measure of information in image

$$Energy = \sum_{j=1}^{n} [P(j)]^2 \qquad 4.6$$

6) Entropy is a measure of randomness and takes low values for smooth images

$$Entropy = - \sum_{j=1}^{n} P(j) \; log_2[P(j)] \qquad 4.7$$

4.2.2 Local Binary Pattern

From work (Ojala, et al., 1996), LBP was used to label the pixel which was implemented thresholding to the neighbourhood of each pixel come along with the output presented as the binary number. In Figure 4.2 is illustrated the work of LBP operation:

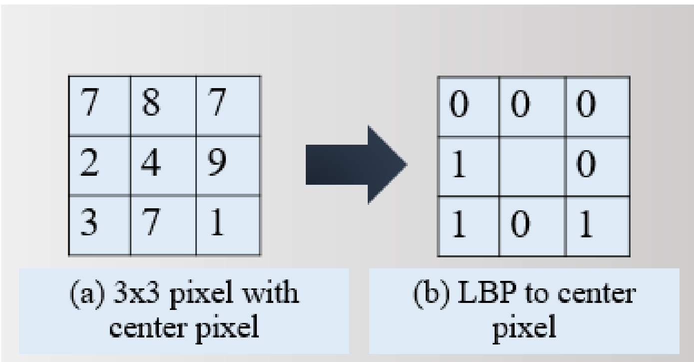


Figure 4.2 LBP Operator

For LBP at pixel (x_c, y_c) can be define as the equation below:

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{p-1} S(i_P - i_c)2^P \quad 4.8$$

where

P presents as the pixel surrounded in the circle neighbourhood.

R presents as the radius of circle.

i_c and i_c present as the grey-level of the centre point.

S(x) presents as the function which S(x) is defined as:

$$S(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases} \quad 4.9$$

In addition, there are some texture descriptors has been proposed as the extension of LBP. In this study, Completed Local Binary (CLBP), Rotated Local Binary Pattern (RLBP) and Local Binary Pattern Histogram Fourier (LBP-HF) are presented below.

4.2.3 Completed Local Binary

In work (Guo, et al., 2010), Completed Local Binary Pattern (CLBP) of a local region can be defined by 2 factors: (i) a center pixel and (ii) a local difference sign-magnitude transform (LDSMT). With the reference to the study purpose, LDSMT has been focused. LDSMT consist of two elements: (i) the difference signs (CLBP S), and (ii) the difference magnitudes (CLBP M).

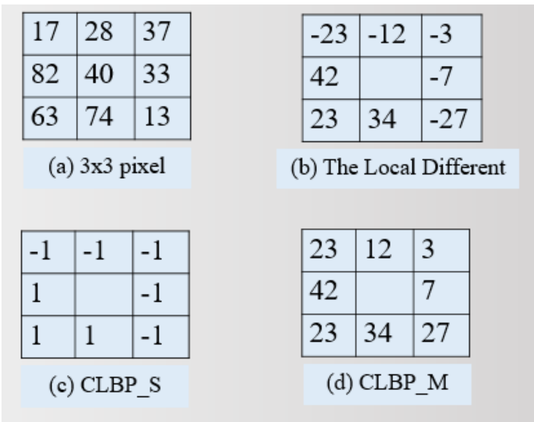


Figure 4.3 The Operation of CLBP

4.2.4 Rotated Local Binary Pattern

Rotated Local Binary Pattern (RLBP) (Mehta and Egiazarian, 2013) or Dominated Rotated Local Binary Pattern (DRLBP) (Mehta and Egiazarian, 2016) is a technique of LBP rotation around the center pixel of object. In order to get the deep understand of RLBP work, the Figure 4.4 is illustrated the process of RLBP work:

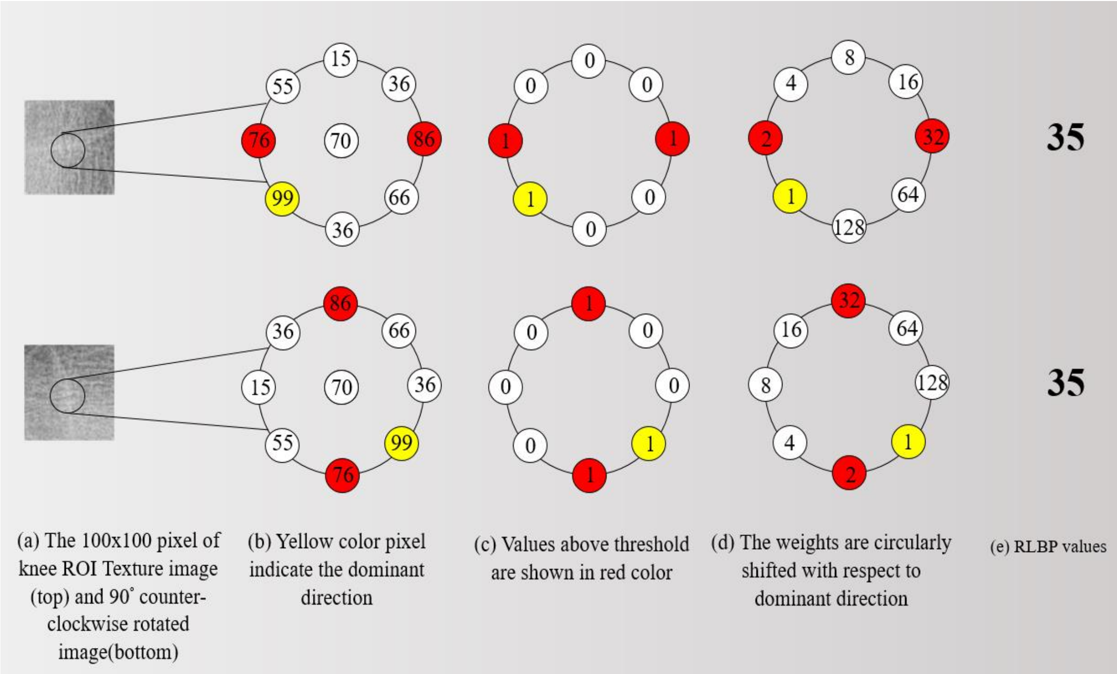


Figure 4.4 The Operation of RLBP

Beside the operation was illustrated in Figure 4.4, RLBP can be defined as the equation below:

$$RLBP_{P,R} = \sum_{P=0}^{p-1} S(i_P - i_c)2^{mod(P-D,P)} \qquad 4.10$$

where

D is dominant direction in a neighbourhood, D can be formed as the equation:

$$D = \underset{P\epsilon(0,1,\dots,P-1)}{argmax} |i_p - i_c|. \qquad 4.11$$

4.2.5 Local Binary Pattern Rotation Invariant

The Local Binary Pattern Rotation Invariant or LBP_{ri} is presented in this sub-section. LBP_{ri} is another extension technique from LBP with rotation invariant feature. Call back to Equation 4.8 LBP expression, LBP operation can produce of output with 2^P different values, mean that the 2^P different binary can be represented by the P pixels of the neighbour set pixel. With the respect to this study, LBP_{ri} have applied to the pixel with 8 neighbour pixels that represented as the equation below:

$$LBP_{ri}(x,y) = \sum_{P=0}^{7} S(i_P - i_x\ , y)2^P \qquad 4.12$$

4.2.6 Local Binary Pattern Histogram Fourier

In this sub-section, the Local Binary pattern Histogram Fourier (LBP-HF) is discussed. LBP-HF considered as a rotation invariant feature descriptor which is an extension from the LBP uniform. LBP-HF is defined by the first computing of histogram non-invariant of LBPs to the whole object or region, further that the invariant feature of the histogram construct rotationally. The main used of LBP-HF is formed by statistic features which is used to calculate the global features of the LBP histogram by Fast Fourier Transform (FFT). In this case, LBP_{ri} consider as the subset of LBP-HF. The implementation of LBP-HF is represented as the Figure 4.5 below:

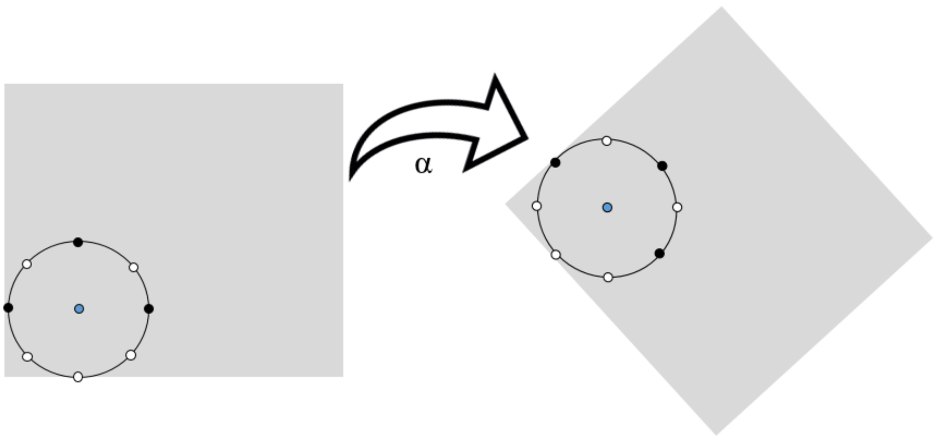


Figure 4.5 The Implementation of LBP-HF

If α = 45°, local binary pattern

10001010 ➔ 00010101

00010101 ➔ 00101010, …,

11111000 ➔ 11110001, …,

At the same time, if α= K * 45°, so, the pattern have to be circularly rotated with K steps.

4.2.7 Local Configuration Pattern

Local Configuration Pattern (LCP) is considered as the rotation invariant image feature description technique. LCP separates image features into two different groups: (i) local structure information and (ii) microscopic configuration or MiC. There are two groups of information architecture include: (i) image configuration and pixel-wise interaction relationship (Guo, et al., 2011). In local structure information is definitely related to the basic of LBP operation, while in microscopic configuration used for searching the feature of microscopic configuration information. In Figure 4.6 is illustrated the implementation of LCP to local structure and MiC concept.

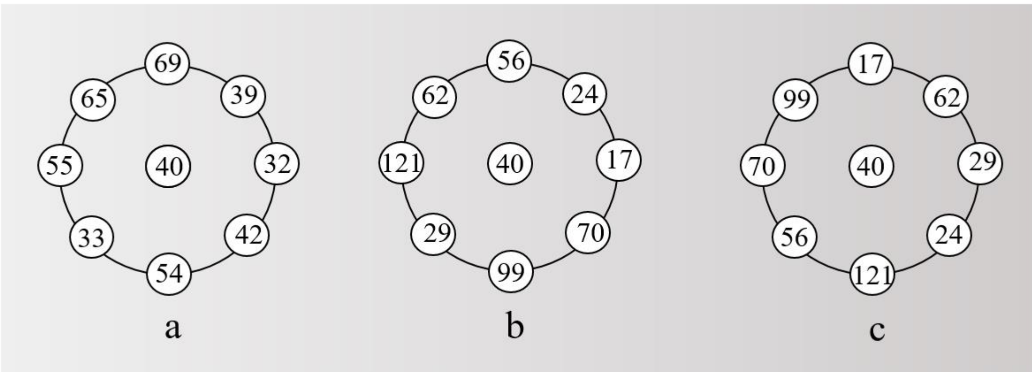


Figure 4.6 The Concept of LCP

Figure 4.6, in Figure 4.6(a) and Figure 4.6 (b) are the same in case of pattern type as LBP (all neighbor pixel compare to center pixel value). On the contrary, Figure 4.6 (a) and Figure 4.6(b) are different in case of LBP with local invariant information, while Figure 4.6(b) and Figure 4.6(c) are the same because of the same value of invariant. For Figure 4.6(b) and Figure 4.6(c) are different if considered in term of MiC. MiC is based on textural properties which can be formed as the equation below:

$$E(a_0, \ldots, a_{P-1}) = |g_c - \sum_{j=1}^{P-1} a_j \, g_j| \quad 4.13$$

where

g_c is the intensity of center pixel values

g_j is the intensity of neighbored pixel values

a_j(j0, …, P-1) are the weighting parameters relation with g_j

$E(a_j)$ is the reconstruction error regarding model parameters of a_j

4.2.8 Local Ternary Pattern

Local Ternary Pattern (LTP) is described in this sub-section. LTP is one of the development techniques from LBP technique that used to analyse the centre pixel of i_c. As noted in Equation 4.8 LBP, S(x) have only two values, but when applied with LCP the S(x) is instead of 3 values function $S_t(u,i_c,t)$ as the following equation:

$$S_t(u, i_c, t) = \begin{cases} 1 \quad if \quad u \geq i_c + t \\ 0 \quad if \quad |u - i_c| < t \\ 0 \quad if \; u \geq i_c + t \end{cases} \quad 4.14$$

4.2.9 Haralick feature

This sub-section presents Haralick feature which is calculated from the Grey-Level Co-occurrence Matrix (GLCM) presents in this sub-section. GLCM is considered as one of the most common techniques to represents the image texture. In this case, Haralick which is measured from the statistic of the basic GLCM is divided into 14 feature as the following:

1) Angular Second Moment (ASM) is used to measures the local uniformity of the grey levels

$$ASM = \sum_{i=1}^{n} \sum_{j=1}^{n} P(i,j)^2 \quad 4.15$$

2) Contrast (standard deviation) is a measure grey level variations between the reference pixel and its neighbour

$$Contrast = \sum_{a=1}^{n} a^2 \; \sum_{i=1}^{n} \sum_{j=1}^{n} P(i,j), |i-j| \qquad 4.16$$

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

$$Correlation = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (ij) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad 4.17$$

4) Variance is the square root value of a grey level variants between the reference pixel and its neighbour measurement

$$\sigma^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} (i-\mu)^2 \, P(i,j) \qquad 4.18$$

5) Inverse Difference Moment (IDM) is similar in concept to the inverse difference feature, but with lower weights for elements that are further from the diagonal:

$$IDM = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{1+(i-j)^2} \, P(i,j) \qquad 4.19$$

6) Sum Average is the average of normalized grey tone image in the spatial domain

$$SumAverage = \sum_{i=2}^{2n} i P_{x+y}(i) \qquad 4.20$$

7) Sum Variance is the variance of normalized grey tone image in the spatial domain

$$Sum\,Variance = \sum_{i=2}^{2n} (i - f_8)^2 P_{x+y}(i) \qquad 4.21$$

8) Sum Entropy (f_8) is a measure of randomness within an image

$$f_8 = -\sum_{i=2}^{2n} P_{x+y}(i) \log[P_{x+y}(i)] \quad 4.22$$

9) Entropy is an indication of the complexity within an image

$$Entropy = -\sum_{i=1}^{n} \sum_{j=1}^{n} P(i,j) \log[P(i,j)] \quad 4.23$$

10) Different Variance is the image variation in the normalized co -occurrence matrix

$$Difference\ Variance = \sum_{i=1}^{n} i^2 P_{x-y}(i) \quad 4.24$$

11) Different Entropy is an indication of the amount of randomness in an image

$$Difference\ Entropy = -\sum_{i=1}^{n} P_{x-y}(i) \log[P_{x-y}(i)] \quad 4.25$$

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

$$IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad 4.26$$

where

HXY is the value of Entropy

HX and HY are the entropy of P_x and P_y

$$\text{HXY1} = -\sum_{i=1}^{n} \sum_{j=1}^{n} P(i,j) \log\{P_x(i) P_y(j)\} \quad 4.27$$

13) Information Measure of Correlation 2 (IMC2) is estimated as follows for symmetric

$$ICM2 = \sqrt{1 - \exp\{-2(HXY2 - HXY)\}} \quad 4.28$$

where

$$HXY2 = -\sum_{i=1}^{n}\sum_{j=1}^{n} P_x(i)P_x(j)\log\{P_x(i)P_y(j)\} \qquad 4.29$$

14) Maximum Correlation Coefficient (MCC) measure of dependence of random variables in image

$$MCC = \sqrt{\sum_{k=1}^{n} \frac{P(i,k)P(j,k)}{P_x(i)P_y(j)}} \qquad 4.30$$

4.2.10 Gabor filter feature

Gabor filter feature descriptor is presented in this sub-section, Gabor is known as one of the texture extraction techniques that used analyse image texture based on specific frequency and specific direction. Gabor filter banks consist of frequencies, orientations and smooth parameters of Gaussian envelope. Gabor filter banks at pixel (x, y) is defined as the equation below:

$$G(x,y) \equiv e^{-\frac{(x-x_0)^2}{2\sigma_x^2}-\frac{(y-y_0)^2}{2\sigma_y^2}} e^{j(\omega_{x_0}x+\omega_{y_0}y)} \qquad 4.31$$

where

ω_{x_0} and ω_{x_0} present as the center frequency of x and y direction

σ_x and σ_y present as the standard of the Gaussian function along x and y direction.

4.3 Feature Selection and Classification

In this section, the feature selection method and learning classifier algorithm which were applied in the proposed framework is presented. As noted in Section 2.5. The five feature selection techniques (CFS, Chi-square, IG, Gain Ratio, and Relief) were applied in this work for OA detection, while CFS is applied for OA stage classification.

For learning classification techniques presented in Sub-section 2.4.3 of Section 2.4. The nine generation classifier presented in Sub-section 2.4.3 were applied in both study of OA detection and OA stage classification.

Finally, the result of each classification go forward to the evaluation process that is discussed in the next section.

4.4 Evaluation

As noted before our research studies was separated into two main objectives: (i) OA detection and (ii) OA stage classification. The detail of evaluation is presented in Sub-section 4.4.1 and 4.4.2 respectively.

4.4.1 Osteoarthritis Detection using texture analysis

In the evaluation of knee-OA detection is presented in this sub-section. The purpose of evaluation was used to produce the evidence that OA condition can be detected efficiency by applying the proposed framework. In this sub-section, the four sets of experiment were evaluated in order to identify the most efficient results for the following objectives:

1) To demonstrate the most appropriate region of interests (ROIs) according to four sub-images presented in Chapter 3 Sub-section 3.4.1

2) To demonstrate the most appropriate texture descriptors in the context of texture descriptors presented in Sub-section 4.2.2

3) To identify the most appropriate feature selection with the respect to feature selection in Section 2.5

4) To identify the most appropriate learning algorithm with the reference to nine learning algorithm discussed in Chapter 2, Sub-section 2.4.3

Ten Cross Validation (TCV) was implemented in this study. To this end, the discussion of each objective, the result were recorded in term of: (i) Area Under Curve (AUC), (ii) Accuracy (AC), (iii) Sensitivity (SN), (iv) Specificity (SP), (v) precision (PR) and (vi) F-measure (FM).

1) Region of Interest

This sub-sub-section describes on the evaluation conducted to compare the result of applying different four ROIs of SET A (As mentioned in Chapter 3 Sub-section 3.4.1): (i) Medial Femur (MF), (ii) Lateral Femur (LF), (iii) Medial Tibia (MT) and (iv) Lateral Tibia (LT). For the experiment LBP descriptor was used with CFS feature selection (LBP and CFS feature selection were used because the reports in Sub-sub-section 4.4.1-II and Sub-sub-section 4.4.1-III, had revealed that this were 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-sub-section 4.4.1-IV).The best performance of each ROI is illustrated in Table 4.1 below (best result indicated in bold font with respect to AUC values):

Table 4.1 The best result of each ROI of OA detection

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	0.895	0.802	0.895	0.802
LT	0.883	0.809	0.809	0.809	0.809	0.809

From Table 4.1 it can be seen that Literal Femur (LF) is the most appropriate one amount of four sub-image of texture analysis for OA detection come with the highest value of AUC of 0.912, while the second highest appropriate went to Medial Femur (MF) with the AUC value of 0.895. It should be suggested that Lateral Femur is first selecting area for texture analysis of OA detection.

2) Texture Descriptor

This sub-sub-section describes on the evaluation conducted to identify the best algorithm for feature extraction (Texture Descriptor) of the sub-image. Ten algorithm of feature extraction 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 was presented in the previous sub-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 4.2 below:

Table 4.2 The best result of each Texture descriptor of OA detection

Texture Descriptor	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
Histogram	0.757	0.695	0.695	0.69	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	0.81	0.81	0.809
LBP_{ri}	0.812	0.771	0.771	0.771	0.771	0.771
LBP-HF	0.773	0.71	0.71	0.717	0.71	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	0.67	0.672	0.662
Gabor	0.883	0.786	0.786	0.786	0.786	0.786

With the reference to Table 4.2, it can be seen that Local Binary Pattern or LBP is considered as the best performance in 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 or 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 4.2, 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.

3) Feature Selection Techniques results

This sub-sub-section describes on the evaluation conducted to determine the best mechanism for dimension reduction of feature vector of each sub-image SET A. Five algorithms of feature selection were included: (i) Correlation-based Feature Selection (CFS), (ii) Chi-Squared, (iii) information gain, (iv) Gain Ration, and (v) Relief feature selection. For the experiments used to compare 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-sub-sections. Again the implementation of Bayesian Network classifier for the same reason as before. The best performance of each feature selection mechanism is illustrated in Table 4.3 below:

Table 4.3 The best result of Feature Selection Techniques of OA detection

Texture Selector	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
CFS	**0.912**	**0.832**	**0.832**	**0.832**	**0.832**	**0.832**
Chi-Squared	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	0.687	0.687	0.687
Relief	0.699	0.679	0.679	0.674	0.681	0.677

In Table 4.3 it can be observed that 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.

4) Classification Algorithm results

This sub-sub-section reports on the evaluation conducted to analyse the best mechanism for generation classifier of learning methods. Nine algorithms of learning methods were considered: (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-sub-sections. The best performance of each learning algorithm is illustrated in Table 4.4 below:

Table 4.4 The best result of Learning Method of OA detection

Learning Method	AUC	AC	SN	SP	PR	FM
C4.5	0.757	0.779	0.779	0.78	0.78	0.779
Binary Split 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
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	0.72	0.715
Logistic Regression	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
Neural Network Backpropagation	0.851	0.771	0.771	0.77	0.771	0.771

With the respect to Table 4.4, it shown that Bayesian Network is the best learning method that can produced 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.

4.4.2 Osteoarthritis stage Classification using texture analysis

The OA stages classification evaluation is presented in this section, the aim of this evaluation was to present the evidence that conclude that OA stage grading is easily to detect by using the proposed approach. With the respected to Section 4.2, the evaluation of section 4.2 shown that CFS mechanism is the most efficient feature selection techniques. Thus in this evaluation, there are three sets of experiment are discussed as the following three objectives:

1) To identify the most appropriate region of interests (ROIs) to the four sub-images, identified ROIs described in Chapter 3, Sub-section 3.4.1

2) To demonstrate the most appropriate feature texture descriptors according to texture descriptors described in subsection 4.2

3) To demonstrate the most appropriate learning methods with respect to nine learning method mentioned in Chapter 2 Sub-section 2.4.3

In all evaluations, Ten Cross-Validation (TCV) was applied and classification recorded in term of: (i) Area Under Curve (AUC), (ii) Accuracy (AC), (iii) Sensitivity (SN), (iv) Specificity (SP), (v) precision (PR) and (vi) F-measure (FM).

1) Region of Interests

This sub-sub-section reports on the evaluation conducted to compare the result of applying different four ROIs of SET A (As mentioned in Chapter 3 Sub-section 3.4.1): (i) Medial Femur (MF), (ii) Lateral Femur (LF), (iii) Medial Tibia (MT) and (iv) Lateral Tibia (LT). For the experiment LBP descriptor was used with CFS feature selection (LBP and CFS feature selection were used because the reports in Sub-sub-section 4.4.2-II and Sub -section 4.4.1 (OA detection study), had revealed that this were 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 stage classification (see Sub-sub-section 4.4.2-III).The best performance of each ROI is illustrated in Table 4.5 below (best result indicated in bold font with respect to AUC values):

Table 4.5 The best result of each ROI of OA stage classification

ROI	AUC	AC	SN	SP	PR	FM
Medial Femur	0.822	0.569	0.569	0.849	0.580	0.553
Lateral Femur	0.816	0.585	0.585	0.849	0.618	0.577
Medial Tibia	**0.871**	**0.654**	**0.654**	**0.907**	**0.691**	**0.658**
Lateral Tibia	0.828	0.592	0.592	0.833	0.635	0.658

From Table 4.5 it can be seen that Medial Tibia (MT) is the best ROI for OA stage detection with the highest value of AUC of 0.871. On the other hands, LF produce the lowest value of AUC value of 0.816. Finally, it can be suggested that medial tibia region is worked well for OA stage detection in case of texture analysis, while lateral femur is considered as the last choice for ROI selection for OA stage detection.

2) Texture Descriptor

This sub-sub-section describes on the evaluation conducted to determine the best algorithm for feature extraction (Texture Descriptor) of the SET A sub-images. Ten algorithm of feature extraction 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 for OA stage classification study, the MT ROI was used as this had been found to produce the best result was presented in the previous sub-sub-section. Again the application of CFS feature selection with Bayesian Network classifier for the same reason as before (the best result of experiment). The best performance of each texture descriptor is illustrated in Table 4.6 below:

Table 4.6 The best result of Texture descriptor of OA stage detection

Texture Descriptor	AUC	AC	SN	SP	PR	FM
Histogram	0.647	0.496	0.496	0.776	0.418	0.408
LBP	**0.871**	**0.654**	**0.654**	**0.907**	**0.678**	**0.658**
CLBP	0.789	0.569	0.569	0.836	0.58	0.553
RLBP	0.817	0.585	0.585	0.856	0.578	0.559
LBP_{ri}	0.682	0.477	0.477	0.815	0.478	0.451
LBP-HF	0.682	0.438	0.438	0.801	0.418	0.416
LTP	0.741	0.508	0.508	0.824	0.532	0.492
LCP	0.747	0.515	0.515	0.834	0.509	0.503
Haralick	0.644	0.454	0.454	0.801	0.431	0.416
Gabor	0.772	0.523	0.523	0.833	0.537	0.514

With respect to Table 4.6 above, it illustrated that LBP the most efficiency texture descriptor in OA stage detection with the highest AUC value of 0.871. The second most efficiency texture descriptor went to RLBP with AUC value of 0.817. On the contrary, Haralick is the lowest performance of texture descriptor with AUC value of 0.644. As a result, it can be suggested that for OA stage detection of X-ray imagery, LBP is the first choice for selecting texture descriptor, while Haralick is the last choice amount of 10 texture descriptors in the study of OA detection.

3) Learning Method

This sub-sub-section describes on the evaluation conducted to analyse the best mechanism for generation classifier of learning methods. Nine algorithms of learning methods were considered: (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 MT ROI and the LBP descriptor were used as this had been found to produce the best result was presented in the previous sub-sub-sections. The best performance of each learning algorithm is illustrated in Table 4.7 below:

Table 4.7 The best result of Learning Method of OA stage detection

Learning Method	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
C4.5	0.628	0.446	0.446	0.798	0.41	0.425
Binary Split Tree	0.659	0.438	0.438	0.832	0.46	0.445
AODE	0.848	0.562	0.562	0.82	0.573	0.518
Bayesian Network	0.858	0.615	0.615	0.834	0.635	0.583
Naïve Bayes	0.854	0.623	0.623	0.827	0.691	0.583
SVM	0.612	0.469	0.469	0.772	0.415	0.353
Logistic Regression	**0.871**	**0.654**	**0.654**	**0.907**	**0.671**	**0.658**
SMO	0.762	0.577	0.577	0.846	0.678	0.646
Neural Network Backpropagation	0.842	0.654	0.654	0.864	0.678	0.646

In Table 4.7, it can be illustrated that Logistic Regression considered as the best result learning methods with the highest AUC value of 0.871. While Bayesian Network preformed as the best second of learning method with the value of AUC of 0.858. In contrast, SVM produced the lowest AUC value with the value of 0.612. It can be concluded that Logistic regression is the best choice for learning method selection for OA stage detection, while SVM is considered as the last choice.

4.5 Discussion

The discussion of knee-OA detection and knee-OA stage classification are presented in this section. With respect to evaluation of the OA classification studies were presented in the previous section comprises of two studies evaluation: (i) knee-OA detection and (ii) knee-OA stage classification.

For knee-OA detection study, the overall classification result of OA detection presented in Sub-section 4.4.1 illustrated that the proposed texture based approach, using ten different texture descriptors for texture analysis to dissimilar four ROIs, performed well to the knee X-ray image dataset. The main finding from the experiment were divided into four sets:

1) Amount of four sub-image, the performance of LF is better than other sub-images performance that can make the classification is more effective. In term of AUC value measure, the report of evaluation found that LF ROI produce AUC value of 0.912 considered as the highest value.

2) The best result of classification performance in term of texture descriptors mechanism for X-ray image dataset, was obtained by applying local binary pattern (LBP), followed by rotated local binary pattern (RLBP), while haralick is the last texture descriptor amount of ten texture descriptor in the experiment.

3) The most appropriate feature selection mechanism in the experiment of the study was obtain by correlation based feature selection (CFS) with the highest AUC value of 0.912, followed by Gain Ration with the AUC value of 0.709, and then three feature selection mechanism: (i) Chi-square, (ii) Information Gain, and (iii) Relief that all these three method produced the same AUC value of 0.699.

4) The best performance of leaning method identified from the reported evaluation were: (i) Bayesian Network, (ii) Logistic Regression, and (iii) Naïve Bayes, which considered as the top three of learning method with the AUC value of 0.912, 0.904, and 0.903 respectively, thus Bayesian network produced a slightly better overall performance than Logistic Regression, while Naïve Bayes almost equal to Logistic regression classifier.

For knee-OA detection study, the overall classification result of OA stage classification presented in Sub-section 4.4.2 identified that the proposed texture based approach, using ten different texture descriptors for texture analysis to different four ROIs, performed well to the knee X-ray image dataset. The main finding from the three sets of experiment conducted were:

1) The best classification performance in term of region of interest or ROI, for knee X-ray image dataset, was obtained by the implementation of Medial Tibia ROI, followed by Lateral Tibia and Medial Tibia with the value of AUC of 0.871, 0.828, and 0.822 respectively.

2) The most appropriate texture descriptors mechanism in the experiment of X-ray image dataset study, was obtained by applying local binary pattern (LBP) with the AUC value of 0.871, followed by rotated local binary pattern (RLBP) with AUC value of 0.817, then completed local binary pattern with the AUC vale of 0.789, while haralick is the last texture descriptor amount of ten texture descriptor in the experiment with the lowest value of AUC of 0.644.

3) The best classification performance in term of leaning method identified from the reported evaluation were: (i) Logistic Regression, (ii) Bayesian Network, and (iii) Naïve Bayes, which considered as the best three learning method with the AUC value of 0.871, 0.858, and 0.854 respectively, thus Logistic Regression produced a slightly better overall performance than Bayesian Network, while Naïve Bayes almost equal to Bayesian Network classifier.

In conclusion, the OA classification studies of the texture based approach in this chapter. With the respect to the discussion of knee-OA detection and knee-OA stage classification study, the main findings of OA classification were comprised into four sets of experiment:

1) In OA detection study the most appropriated ROI referred to the (Lateral Femur) LF Region of Interest (ROI), while in Osteoarthritis (OA) stage detection the most appropriated ROI went to (Medial Tibia) MT ROI. Thus, it can be suggested that for OA detection the Femur region is more important than Tibia region.

On the other hand, for OA stage detection the Tibia region should be selected at the first choice, due to the performance is better than Femur region.

2) For the second point of the conclusion, the best performance of Texture descriptor of both OA detection and OA stage detection went to (Local Binary Pattern) LBP texture descriptor that gave the best result of AUC and AC and the second best performance of texture descriptor went to Rotated Local Binary Pattern or RLBP.

3) In the third pint of the conclusion, the most appropriated feature selection technique amount of five methods in Osteoarthritis (OA) detection was performance by Correlation-based Feature Selection or CFS feature selection techniques, while in OA stage detection study had applied only CFS that got the advance study from OA detection Study.

4) The best classification result of OA detection was performance by classifier generation obtained by Bayesian Network, follow by Logistic Regression, and then Naïve Bayes classifier. In contrast, in OA stage detection was obtained by Logistic Regression, Bayesian Network, and Naïve Bayes method. In conclusion, for both study of OA detection and OA stage detection best classification result in term of classifier generation followed by the best three classifier: (i) Bayesian Network, (ii) Logistic Regression, and (iii) Naïve Bayes.

Finally, there are four main points of the both study (OA detection and OA stages classification). As a result, for OA detection LF region, LBP Texture Descriptor, CFS feature selection technique and Bayesian network can be combined each other to produce the best performance of OA detection study. For OA stages classification, the MT region, LBP texture descriptor, CFS feature selection technique, and Logistic Regression combine each other to produce the best performance of OA stage classification.

4.5 Summary

To summary, the chapter presented the texture based approach on knee OA detection and knee OA stage detection study. For OA detection study base on texture was presented in Section 4.2 and OA stage detection was presented in Section 4.3. Base on the reported of each section shown that: (i) Lateral Femur is the best suitable ROI for OA detection, while Medial Tibia is the most efficient ROI for OA stage detection, (ii) LBP is the best texture descriptor for both study, (iii) CFS is play the most significant feature selection in OA detection, while OA stage use only CFS, and (iii) Bayesian Network, Logistic Regression, and Naïve Bayes is the most three efficient learning method for both study.

CHAPTER 5

OSTEAOARTHRITIS CLASSIFICATION USING KNEE X-RAY IMAGERY: GRAPH-BASED APPROACH

5.1 Introduction

The introduction of graph based approach applying for knee-OA detection and knee-OA stage classification are considered in this section. The major objective of the knee-OA detection study is to classify OA and normal control, while knee-OA stage classification work aims to classify the stage of knee-OA with 128 medical X-ray images presented in Chapter 3. The promoted idea of this section is illustrated the nature of each Whole knee and knee joint space X-ray image which is the analysing of the bone shape (shape analysis), using graph based representation. In term of training data, the graph based present of the dataset segmentation as whole knee segmentation and knee joint space segmentation. This training can then be applied to build a graph based classifier that can be used to analyses OA classification of image according to the nature of proposed graph structure representation.

To be more specific, the image decomposition approach that used for bone shape analysis is discussed where by the whole knee and knee joint space sub-image are presented using quadtree decomposition, both whole knee and joint space sub-image were presented in Section 3.4 of Chapter 3. In order to get the joint space more clear, the Otsu was applied to joint space sub-image. Thus, the three subset datasets for study. Once each set of sub-image has been fine segmented the next stage of the data preparation phase is to translate the segmented sub-image pixel dataset into a form of suitable for the application of classifier. The data translation need to be conducted in a way of better information is selected while in the intervening time ensuring that the representation is better enough to enable for effective further processing. Typically, a quadtree representation (one per each sub-image). In case of medical image, work (Dua, et al., 2010 and Khalili, et al., 2013) have been applied quadtree for the proposed study of medical image classification and segmentation. On the other hands, the quadtree representation does not depend on itself to ready incorporation with reference to learning methods. In order to do this, the subgraph mining was applied to the quadtree data to identify frequently subgraph which frequently occurring pattern across data that can be considered as feature in term of a feature vector representation. A proposed framework of graph based approach is given in Figure 5.1bellow:

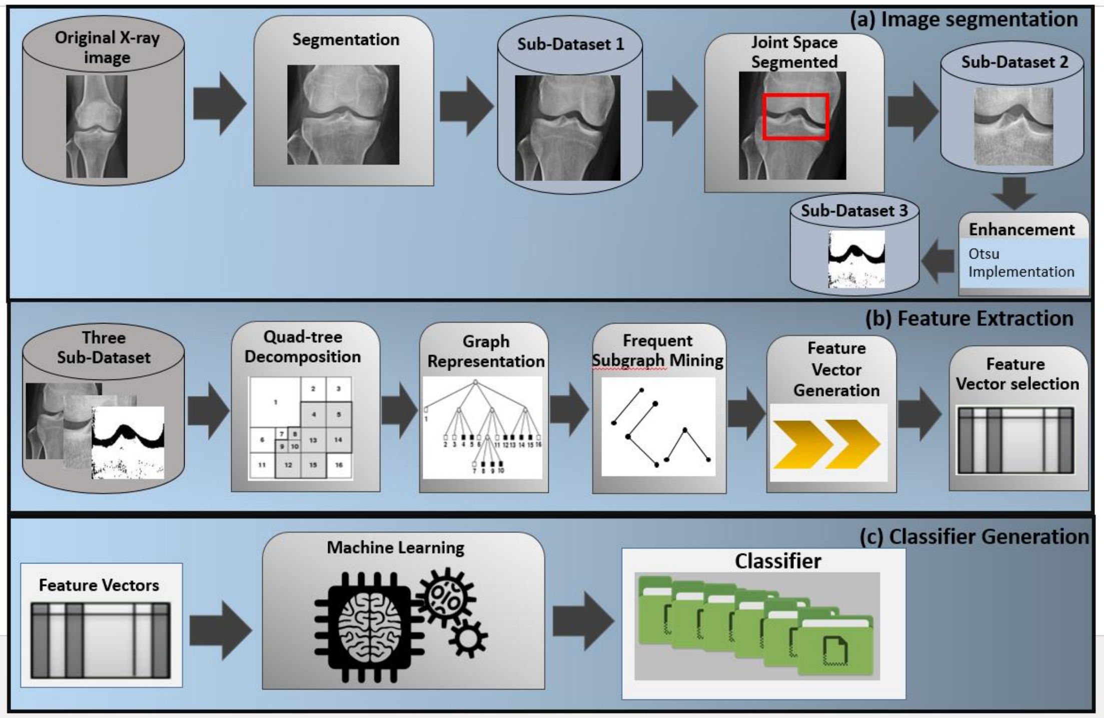


Figure 5.1 The Graph-Based of OA classification framework

From the Figure 5.1, it can be observed that graph based of OA classification comprises of three main processes: (a) Image segmentation, (b) Feature Extraction, and (c) Classifier Generation. For the process of image segmentation and enhancement considered as the first process of the framework was presented in Chapter 3 and will thus not be considered more in this chapter. A collection of Data SET B mentioned in Chapter 3 Section 3.4 was used. There were three sets of image that obtained from the segmentation and enhancement processes: (i) the whole knee sub-image, (ii) the joint space sub-image, and (iii) the enhancement of joint space sub-image (Otsu implementation). Once a dataset of each sub-image has been segmented, the next process refers to the feature extraction process, which used to translate the segmented image pixel data into an appropriated form suitable for classier generation (the third process of the study framework). The third process of the framework is little further consideration in this section.

In the feature extraction as presented in Figure 5.1 contained a number of sub-processes. The major idea of the processing is to apply the graph based approach using quadtree decomposition to each sub-image dataset. In this case, the group of subgraph that frequently happen in the data can be established. Thus, a feature vector representation of the form used by classifiers generation. The sub-process that create the feature extraction process are comprised of five sub-processes: (i) quadtree decomposition, (ii) graph/tree representation, (iii) frequent subgraph mining, (iv) feature vector generation, and (v) feature selection.

The rest of this chapter is organised as follow: the fundamental idea of quadtree decomposition is presented in Section 5.2, and Section 5.3 reports the tree/graph representation. The discussion of frequent subgraph mining is illustrated in Section 5.4, while the Section 5.5 presents the feature selection and classification. The evaluation of the chapter is illustrated in Section 5.6, for knee-OA detection evaluation is presented in Sub-section 5.6.1 and Sub-section 5.6.2 illustrates the knee-OA stage classification evaluated. The discussion of OA classification work is presented in Section 5.7. Finally, the chapter summary is presented in Section 5.8.

5.2 Quadtree Decomposition

In this section the quadtree decomposition is presented. Image decomposition considered as the methodology for "factorising an input image in to a group of component" (Chen and Koltun, 2013). Image decomposition applications has been applied in various applications: (i) image classification, (ii) Image segmentation,

(iii) image recognition, (iv) image fusion, (v) computer vision, and (vi) motion estimation. The methodology of image decomposition comprises of (i) quadtree, (ii) wavelet, (iii) scale space, and (iv) pyramid (both Gaussian and Laplacian pyramid). To be more evidence of the methodology of image composition in term of medical image, work (Mittal, et al., 2015) have been applied wavelet parameter to MRI and CT images classification, in the study (Sahu, et al., 2014) have proposed the Laplacian pyramid for image fusion with the application of CT and MRI image.

In the context of digital image, the four method of image decomposition (quadtree, wavelet, scale space, pyramid) are presented above were proposed in works (Wang, et al., 2015 and Samet, 1984). In this work refer to the applying of graph-based approach to each sub-image. Quadtree decomposition is known as a technique of a hierarchical approach of decomposing in image that naturally depend on quadtree data structure. The most well-known application of quadtree decomposition is applie in region base quadtree, where each level of the quadtree decomposition decompose image in to quadrant (four equal regions) (Samet, 1990). One problem that is considered as the most common issue of quadtree decomposition is the stopping of the level to be adopted. In order to solve this problem, the maximum level of decomposition is expressed.

In the context of the knee X-ray image data considered in this chapter a quadtree decomposition of region of interest was applied. Figure 5.2 presents the example of quadtree decomposition process. Figure 5.2(a) illustrates the original image, Figure 5.2(b) presents a 23 x 23 image binary array where ‘1s’ are pixel insight the region, while ‘0s” are the pixels outside the region. Figure 5.2(c) illustrates the result of applying Quadtree decomposition to the region. In the Quadtree, the whole image is represented as the root node, while the immediate child nodes of the root nodes each refers as a region quadrant. Figure 5.2 illustrates the processes of quadtree decomposition, the process stops when the process arrives at the homogenous region (region consist of all black pixels or all white pixels). The quadtree representation of figure 5.2(c) is illustrated in Figure 5.2(d).

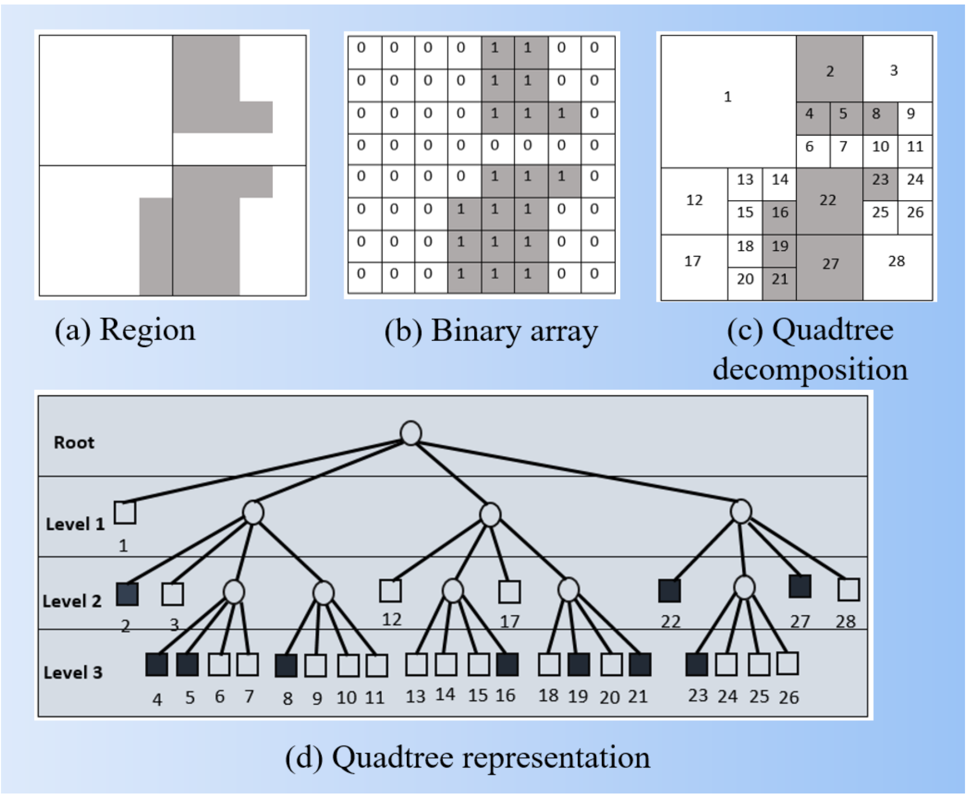


Figure 5.2 The Quadtree decomposition

In the context of X-ray image used in this research, the identified ROI was resized into 128 x 128 pixel square image (the sub image is manually cropping). In Figure 5.3 is illustrated the process of quadtree decomposition to the whole knee sub-image (note that there are three sub-image dataset were used in this chapter, the author picked up the whole knee sub image as an example). Figure 5.3(a) shows the original of whole knee sub image, while Figure 5.3(b) illustrates the quadtree decomposition of Figure 5.3(a).

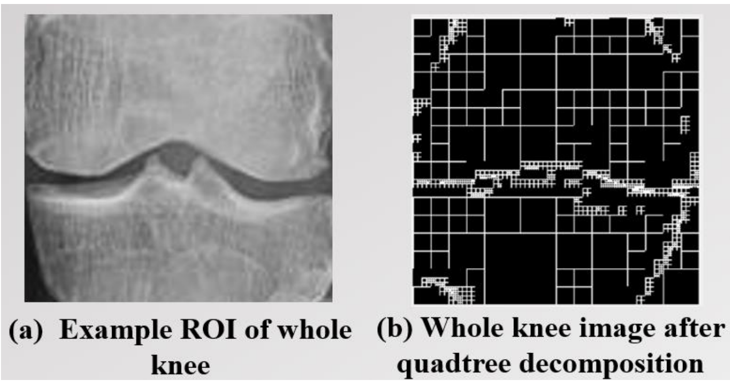


Figure 5.3 The example of Quadtree Decomposition to Whole knee sub-image

5.3 Tree Representation

Once the segmented sub-image segmented have been decomposed and store in quadtree format, as described in 5.2, The sub-mage was resized to *128 x 128* size, in this case eight labels were derived where each of label describe a range of 32 consecutive intensity value. Figure 5.4 is illustrated the example of quadtree representation where the root (top level node) represent as the whole knee sub image. Figure 5.4 the next level of the root level is considered as the Level 1 is the root node's immediate child nodes, and so on. From the root node or the parent node separates the edges into a set of identifiers 1, 2, 3, and 4 illustrating the NW, NE, SW, and SE. In Figure 5.4 illustrates that the number in square brackets alongside each node is a unique node identifier derived according to the decomposition.

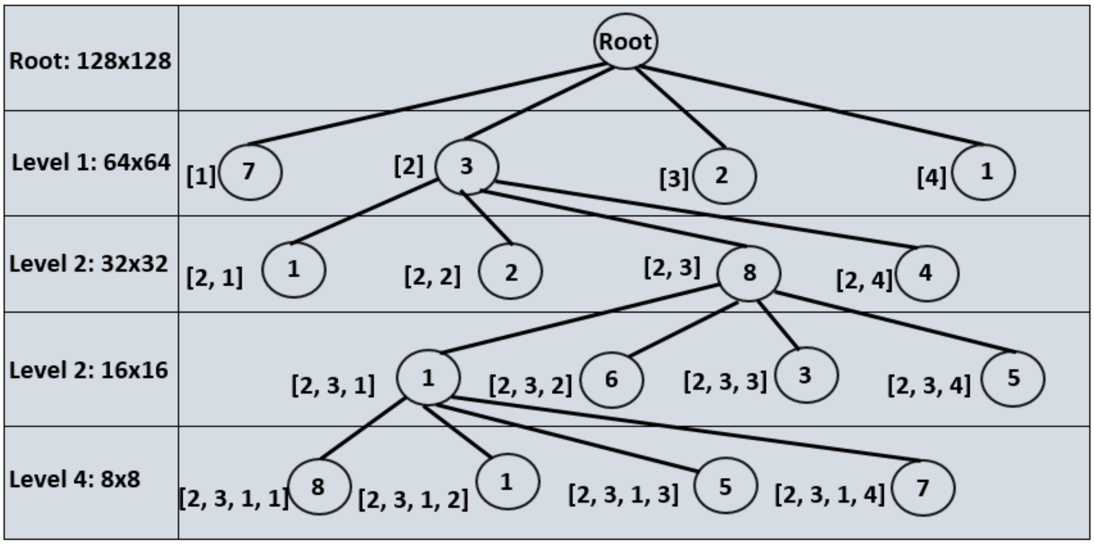


Figure 5.4 The Quadtree representation

5.4 Frequent Subgraph Mining

Frequent Subgraph Mining (FSM) is a technique of hierarchical decomposition image. The fundamental of FSM is to identify a group of feature which can used to create a feature vector representation.

In other words, FSM is well-known graph mining technique, FSM is the process of indicating the hidden information in graph data. Graph representation which is considered as a popular techniques of graph mining, are widely applied and work as the powerful and flexible approach for representing or modelling entities include chemical compounds, protein structure, circuit, biological network, work-flows, social networks, world wide web information, xml document and image data (Han and Kamber, 2011 and Olmos and Gonzalez, 2008). For work chemical informatics, computer vision, video indexing and text retrieval were included in graph mining application (Haddad and Kheddouci, 2007; Lai, et al., 2015; Riaz and Ali, 2011 and Xu, and Wang, 2011).

In this chapter, the graph mining which from FSM is the most appropriate technique to apply. Frequent subgraph mining is a famous graph mining technique which is mention in this study. Frequent subgraph is a technique to discover the graph that happen frequently, frequent subgraph may be used to: discriminate between different sets of graphs, characterise graph set, classify and cluster graphs, and facilitate similarity search in graph database. The work (Han, J., and Kamber, 2011) and work (Huan, et al., 2003) have applied FSM for chemical analysis. FSM can be used in order to a collection of graphs or one single large graph, in the case of the this study, a collection of graphs representing the three different sub-image dataset (whole knee, knee joint space, and Ostus thresholding to knee joint space sub-image dataset) segmented from the knee original image. Hence given a graph dataset $D=\{G_0, G_1, \ldots, G_n\}$, *support(g)* present the number of graphs (in D) in which a subgraph g exists. A subgraph g can be considered as the frequent if *support(g)* $\geq \sigma$, where σ is a minimum support threshold.

The isomorphism testing is considered as the main component of FSM algorithm, it is the process of reviewing whether a subgraph g_i is identical to a subgrap

g_j. Isomorphism is needed with reference to candidate and support counting. Isomorphism is the major computational overhead associated with FSM. The majority FSM algorithm seek to limit the amount of isomorphism that is require. Apriori-based Graph Mining (AGM) algorithm is the given example proposed in work (Inokuchi, et al., 2000), then develop to be the Frequent Subgraph Mining (FSM) presented in work (Kuramochi and Karypis, 2001) which the FSM based on the idea of using what the author refer to as the “adjacent representation” of graph and an “edge-growing” strategy. In work (Agrawal and Srikant, 1994) presented both of AGM and FSM take advantage of the Apriori mechanism in case of frequent item set mining of tabular data, example, if a *k* edge subgraph is not frequent none of its *k+1* edge subgraph will be frequent. The SFM of Apriori style comprises into three steps: (i) candidate generation, (ii) support counting, and (iii) graph pruning based on the σ threshold value. In the time of candidate generation *k+1* edge candidate subgraph are generated from the frequent *k* edge subgraph identified on the previous iteration, a process called as *subgraph growing*.

Algorithm 1: The Process of Frequent Subgraph Mining

INPUT G = $\{G_0, G_1, \ldots, G_n\}$, σ = threshold;

OUTPUT S = $\{S_0, S_1, \ldots, S_n\}$;

S = null;

k = 1;

C_k = all one edge candidate subgraph in G;

loop

  L = set of occurrence count for each $G_i \epsilon C_k$ obtained using an isomorphism process, with one to one corresponding with C_k;

  F = set of frequent subgraph in C_k, where for each $g_i \epsilon C_k$ $1 \epsilon L < \sigma$;

  S = S ∪ F;

  k++

  C_k = set of k edge subgraph extended from F using right most extension

  if (k==null) then exit

end loop

The Algorithm 1 above describe the frequent subgraph process. In the algorithm, the input comprise of two variables: (i) a collection of graph (each graph represent an image) denote by G and (ii) the threshold value σ. A collection of frequent subgraph denoted by S is the outcome of the output. There are three parameter are used for the begins of the process: (i) the set of frequent subgraph is initially an empty set, (ii) the counter k is defined as 1, and (iii) the set of one edge candidate subgrap is defined as C_k. In the algorithm, there is one which is start from line 6 to line 13 through the following steps: (i) determine the occurrence count for each subgraph $G_i \in C_k$ using an isomorphism process, (ii) compare the occurrence count of each subgraph $G_i \in G_k$ with the σ and add those G_i whose occurrence count is greater than σ to the set F. (iii) add the set F to the set S, (iv) increment k and generate the next size of candidate sets C_k from F, and (v) do the loop of the process again. The loop does until no more candidate can be generated (C_k =0).

In the study, graph representation have been applied to the three subimage of knee X-ray image contain of two main advantages: (i) the first advantage of the representation is the "boundary problem" where object of interest may be located at the intersection of a decomposition and (ii) the quadtree representation is not directly suited to use with classifier generation and subsequent usage of the generated classifier. Thus, the FSM have been applied for the purpose of this study to be considered as the feature within feature vector representation that is suitable to apply with classifier generation.

When the set of frequently has occurred, the frequently can be arranged into a feature vector representation, for each feature vector indicates the presence of a particular subgraph with the reference to each knee subimage as illustrated in Table 5.1. With the respect to the table each row refer to the individual subimage image knee data of each sub dataset (record from 1 to m), and the columns individual frequent subgraph present by set $\{S_1, S_2, \ldots, S_n\}$. In the table row, the value 0 represent the absence and the value 1 indicate the presence of the associated subgraph for the record. The output of this process is the feature vector which can be apply for both classifier generation and the future usage of the generated classifier.

Table 5.1 The example of Feature Vector of Knee X-ray Image

Vector	S_1	S_2	S_3	S_4	S_5	….	S_n
1	1	1	0	0	1	…	1
2	0	1	1	1	0	…	1
3	1	0	0	1	1	…	0
…	…	…	…	…	…	…	…
m	1	0	1	0	1	…	1

5.5 Feature Selection and Classification

Once feature vector generation was completed from the previous process, then the classification model generation could be considered. However, in order to generate the classification model, the input data need to be discretised (range). The challenge of the feature selection process is the large number of feature (subgraphs) identified. Feature selection have been applied to reduce the feature dimensionality whereby only highly discriminative features were retained. In this study, the Correlation-based Feature Selection (CFS) have been applied to feature evaluation measure for scoring feature. On the finishing of feature selection process, each subimage of each dataset was describe in term of a reduce number of features.

After the feature selection process was done, the next process is the classification process. The extensive evaluation was conducted so as to test operation of the different parameters and their variation, in this study only discuss the most significant results. With the reference to the evaluation in the next sub-section, two well-known method of Bayesian Network and Naïve Bayes to the three different subdataset where each dataset was considered as *Dataset* 1 (whole knee sub-image dataset), *Dataset* 2 (knee joint space sub-image dataset), and *Dataset* 3 (the implementation of Otsu to knee joint space sub-image dataset). For Bayesian Network and Naïve Bayes are taken from the Waikato Environment for Knowledge Analysis (WEKA) machine learning workbench.

5.6 Evaluation

As noted before the research study was divided into two main objectives: (i) OA detection and (ii) OA stage Classification. The detail of evaluation is described in Sub-section 5.6.1 and 5.6.2 respectively.

5.6.1 Osteoarthritis Detection using Quadtree analysis

In the evaluation of knee-OA detection is described in this sub-section. The purpose of evaluation was used to produce the evidence that OA condition can be detected efficiency by applying the graph-based framework (shape analysis). In this evaluation, the three sets of experiment were evaluated in order to identify the most efficient results for the following objectives:

1) A set of experiments to identify the most appropriate data set for use with respect to OA detection.

2) A set of experiments to analyse the most appropriate support threshold according to graph mining.

3) A set of experiments to determine the most appropriate classification learning methods for OA detection.

1) Data Set

This sub-sub-section reports on the evaluation conducted to compare the result of applying three different sub-images of SET B (As mentioned in Chapter 3 Section 3.4), each sub-image is illustrated in Figure 5.5.

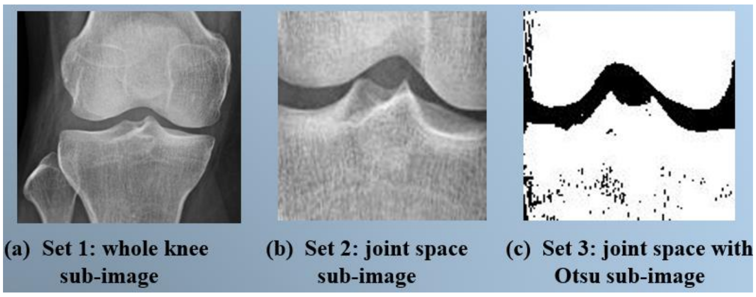


Figure 5.5 Example of SET B sub-images

For the experiment the support threshold value of σ =10 was used with CFS feature selection (support threshold and CFS feature selection were used because the reports in Sub-sub-section 5.6.1-II and the previous knee-OA detection study, had revealed that this were appropriate texture descriptor and feature selection, respectively) and a Naïve Bayes classifier method as these had been found to work well in the context of knee-OA stage classification (see Sub-sub-section 5.6.1-III).The best performance of each image sets is illustrated in Table 5.2 below (best result indicated in bold font with respect to AUC values):

Table 5.2 The best result of each data set of sub-image of OA detection

Data set	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
Dataset 1	0.895	0.781	0.781	0.786	0.789	0.780
Dataset 2	0.88	0.781	0.781	0.777	0.785	0.78
***Dataset* 3**	**0.917**	**0.828**	**0.828**	**0.826**	**0.829**	**0.828**

With the reference to table 5.2 reported that the applying of Otsu method to the knee joint space sub-image (*Dataset* 3 sub-image) produce the best record of the research experiment with the AUC record of 0.917. In the same time, the whole knee segmented image (*Dataset* 1 sub-image) produced the second best result with the AUC value of 0.895. It should be suggested that the applying of Otsu to knee joint space segmented image works perfectly on knee OA detection in this research study.

2) Subgraph Mining

This sub-sub-section presents on the evaluation conducted to compare the support threshold value (σ) for subgraph mining. Five different value of support threshold value were applied in subgraph mining: (i) σ =10, (ii) σ=20, (iii) σ=30, (iv) σ=40 and (v) σ=50. For the experiments used to compare these five support threshold the *Dataset* 3 sub-image was used as this had been found to produce the best result was presented in the previous sub-sub-section. Again CFS feature selection was adopted together with Naïve Bayes classifier for the same reason as before. In addition the number of features for each support threshold in presented in Table 5.3.

Table 5.3 The best result of support threshold value of OA detection

σ	***Dataset* 1**	***Dataset* 2**	***Dataset* 3**
σ= 10	373	389	991
σ = 20	117	99	235
σ= 30	63	48	112
σ= 40	33	26	61
σ= 50	13	14	30

From the Table 5.3 it can be observed that the number of attribute or feature is decreased while the support threshold is increasing.

The best performance of each support threshold for knee-OA stage classification study is reported in Table 5.4 below:

Table 5.4 The best result of support threshold value of OA detection

σ	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
σ= 10	**0.917**	**0.828**	**0.828**	**0.826**	**0.829**	**0.828**
σ = 20	0.861	0.797	0.797	0.797	0.797	0.797
σ= 30	0.823	0.75	0.75	0.749	0.75	0.75
σ= 40	0.806	0.727	0.727	0.727	0.727	0.727
σ= 50	0.74	0.703	0.703	0.704	0.704	0.703

From Table 5.4 it can be observed that the best result of the experiment from the support threshold value of σ =10 with the AUC value of 0.917, then the second best result and the third best result are indicated by the value of σ =20 and σ=30 with the AUC vale of 0.861 and 0.823 respectively. It should be noted that knee OA detection in term of graph based approach of applying quadtree the best suitable support threshold value of 10.

3) Learning Method

This sub-sub-section presents on the evaluation conducted to determine best mechanism for generation classifier of learning methods. Nine of learning algorithms were considered: (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 *Dataset* 3 sub-image set and support threshold value σ =10 were used as this and applied with CFS feature selection had been found to produce the best result was presented in the previous sub-sub-sections. The best performance of each learning algorithm is reported in Table 5.5 below:

Table 5.5 The best result of learning mechanism for OA detection

Learning Methods	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
Decision Tree	0.731	0.703	0.703	0.704	0.704	0.703
Binary Split Tree	0.731	0.703	0.703	0.704	0.704	0.703
AODE	0.916	0.828	0.828	0.826	0.829	0.828
Bayesian Network	0.915	0.828	0.828	0.826	0.829	0.828
Naïve Bayes	**0.917**	**0.828**	**0.828**	**0.826**	**0.829**	**0.828**
SVM	0.81	0.813	0.813	0.807	0.819	0.811
Logistic Regression	0.832	0.758	0.758	0.758	0.759	0.758
SMO	0.805	0.805	0.805	0.806	0.806	0.805
Back propagation	0.873	0.805	0.805	0.804	0.805	0.805

From Table 5.5 it can be determined that Naïve Bayes is the best learning method that can produced the highest value of AUC with the value of 0.917, while the second best learning method went to AODE with the AUC value of 0.16 and the Bayesian Network is the one of the top three algorithm with the AUC value of 0.915. In contrast, decision tree and binary split tree are the lowest learning method for selection in case of OA detection due to the production of AUC value of 0.731 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 Naïve Bayes to *Dataset* 3 sub-image, and CFS feature selection approach perform well for OA detection.

5.6.2 Osteoarthritis Stage Classification using quadtree analysis

In the evaluation of OA stage classification is reported in this sub-section. The purpose of evaluation was used to provide the evidence that OA stage can be classified efficiency by applying the proposed framework (shape analysis). In this sub-section, the different three sets of experiment were evaluated in order to identify the most efficient results for the following objectives:

1) A set of experiments to identify the most appropriate data set for use with respect to OA stage classification.

2) A set of experiments to analyse the most appropriate support threshold according to graph mining for OA stage classification.

3) A set of experiments to determine the most appropriate classification learning methods for OA stage classification.

1) Data Set

This sub-sub-section reports on the evaluation conducted to compare the result of applying three different sub-images of SET B (As mentioned in Chapter 3 Section 3.4): (i) whole knee sub-image, (ii) joint space sub-image and (iii) joint space with Otsu (each sub-image was presented in the previous sub-section). For the experiment the support threshold value of σ =10 was used with CFS feature selection (support threshold for subgraph mining and CFS feature selection were used because the reports in Sub-sub-section 5.6.2-II and the previous knee-OA detection study, had revealed that this were appropriate texture descriptor and feature selection, respectively) and a Bayesian classifier method as these had been found to work well in the context of knee-OA stage classification (see Sub-sub-section 5.6.2-III).The best result of each image set is illustrated in Table 5.6 below (best result indicated in bold font with respect to AUC values).

Table 5.6 The best result of each data set of sub-image

Data set	AUC	AC	SN	SP	PR	FM
Dataset 1	0.761	0.523	0.523	0.777	0.624	0.447
Dataset 2	0.782	0.539	0.539	0.785	0.635	0.46
***Dataset* 3**	**0.819**	**0.539**	**0.539**	**0.774**	**0.551**	**0.44**

From table 5.6 it can be seen that the applying of Otsu's method to the knee joint space sub-image (*Dataset* 3 sub-image) produce the best record of the Knee-OA stages classification research experiment with the AUC record of 0.819. For the best second appropriate sub-image went to the knee joint space segmented image (*Dataset* 2 sub-image) produced the second best result with the AUC value of 0.782. It should be concluded that the applying of Otsu to knee joint space segmented image works well on knee OA stage detection study.

2) Subgraph Mining

This sub-sub-section report on the evaluation conducted to compare the support threshold value (σ) for subgraph mining of knee-OA stage classification. Five different value of support threshold value were applied in subgraph mining: (i) σ =10, (ii) σ=20, (iii) σ=30, (iv) σ=40 and (v) σ=50. For the experiments used to compare these five support threshold the *Dataset* 3 sub-image (joint space with Otsu thresholding) was used as this had been found to produce the best result was presented in the previous sub-sub-section. Again CFS feature selection was adopted together with Naïve Bayes classifier for the same reason as before.In addition the number of features for each support threshold of knee-OA stage classification study is presented in Table 5.7.

Table 5.7 The number of feature for each dataset with support threshold value

σ	***Dataset* 1**	***Dataset* 2**	***Dataset* 3**
σ= 10	346	413	1046
σ = 20	113	99	278
σ= 30	52	61	132
σ= 40	31	32	70
σ= 50	12	18	36

From the Table 5.7 it can be observed that the number of attribute or feature is decreased while the support threshold is increasing.

The best performance of each support threshold for subgraph mining is reported in Table 5.8 below:

Table 5.8 The best result of support threshold value of knee-OA stage classification

σ	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
σ= 10	**0.819**	**0.539**	**0.539**	**0.774**	**0.551**	**0.44**
σ = 20	0.751	0.484	0.484	0.774	0.401	0.442
σ= 30	0.753	0.445	0.445	0.735	0.483	0.473
σ= 40	0.723	0.461	0.461	0.742	0.472	0.472
σ= 50	0.679	0.445	0.445	0.804	0.408	0.401

From Table 5.8 it can be reported that the best result of the experiment from the support threshold value of σ =10 with the AUC value of 0.819, then the second best result and the third best result are indicated by the value of σ =30 and σ=20 with the AUC vale of 0.753 and 0.751 respectively. It should be suggested that knee OA stages classification study in term of graph based approach of applying quadtree the best suitable support threshold value of 10.

3) Learning Method

This sub-sub-section describes on the evaluation conducted to analyse the best algorithm for generation classifier of learning methods. Nine of learning mechanism were considered: (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) BackPropagation Neural Network. For the experiments used to compare these nine mechanisms the *Dataset* 3 sub-image set and support threshold value σ =10 were used as this and applied with CFS feature selection had been found to produce the best result was presented in the previous sub-sub-sections. The best performance of each learning algorithm is reported in Table 5.9 below:

Table 5.9 The best result of learning mechanism for OA stage classification

Learning Methods	**AUC**	**AC**	**SN**	**SP**	**PR**	**FM**
Decision Tree	0.626	0.391	0.391	0.786	0.405	0.393
Binary Split Tree	0.626	0.391	0.391	0.786	0.405	0.393
AODE	0.803	0.469	0.469	0.734	0.319	0.37
Bayesian Network	**0.819**	**0.539**	**0.539**	**0.774**	**0.551**	**0.44**
Naïve Bayes	0.801	0.484	0.484	0.743	0.328	0.383
SVM	0.603	0.453	0.453	0.752	0.308	0.358
Logistic Regression	0.659	0.438	0.438	0.833	0.472	0.472
SMO	0.79	0.555	0.555	0.834	0.567	0.555
Backpropagation	0.701	0.414	0.414	0.81	0.422	0.416

From Table 5.9 it can be observed that Bayesian Network classifier is the best learning method that can produced the highest value of AUC with the value of 0.819, while the second best learning method went to AODE with the AUC value of 0.803 and the Naïve Bayes is best third algorithm with the AUC value of 0.801. In contrast, SVM is the lowest learning method in case of OA stage classification due to the production of AUC value of 0.603 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 *Dataset* 3 sub-image, and CFS feature selection approach perform well for OA detection.

5.7 Discussion

The discussion of knee-OA detection and knee-OA stage classification using quadtree analysis are presented in this section. As noted to the previous section of this chapter, there were 2 study of OA classification studies: (i) knee-OA detection and (ii) knee-OA stage classification.

As noted in the knee-OA detection result in the previous section illustrated that the proposed of quadtree applying in term of graph-base approach, using quadtree analysis to three different subimage, performed well to the knee X-ray image dataset. The main finding from the knee-OA detection study experiment were divided into three sets:

1) Amount of five support threshold value (σ), the performance of σ value of 10 (σ=10) is better than other support threshold values performance that can make the classification is more effective. In term of AUC value measure, the report of evaluation found that σ=10 value produce AUC value of 0.916 considered as the highest value.

2) The most appropriate sub-image of ROIs experiment of the study was obtain by the sub-image of the applying Otsu to knee joint space sub-image (Set 3 sub-image) with the highest AUC value of 0.916, followed by the whole knee segmented image (Set 1 sub-image) with the AUC value of 0.877.

3) The best performance of leaning method identified from the reported evaluation were: Naïve Bayes and Bayesian Network, which considered as the top performance of learning method with the AUC value of 0.916 and 0.915 respectively, thus Naïve Bayes produced a slightly better overall performance than Bayesian Network.

With knee-OA stage classification presented in the Sub-section 5.6.2 of this chapter, The overall classification result of OA stages detection presented in the previous section, section 6.3.6 illustrated that the proposed of graph-based approach, using quadtree analysis to three different sub-image, and the proposed work smoothly to the knee X-ray image dataset. The main finding from the experiment were divided into three sets:

1) Amount of five support threshold value (σ), for example σ=10, 20, 30, 40 and 50, the performance of σ value of 10 (σ=10) is better than other support threshold values performance that can make the classification is more effective. In term of AUC value measure, the report of evaluation found that σ=10 value produce AUC value of 0.851 considered as the highest value.

2) The most appropriate sub-image of ROIs experiment of the knee OA stages classification study was obtain by the sub-image of the applying Otsu to knee joint space sub-image (Set 3 sub-image) with the highest AUC value of 0.851, followed by the whole knee segmented image (Algorithm 1 sub-image) with the AUC value of 0.803.

3) The best performance of leaning method identified from the reported evaluation were: Naïve Bayes and Bayesian Network, which considered as the top performance of learning method with the AUC value of 0.851and 0.845 respectively, thus Naïve Bayes produced a slightly better overall performance than Bayesian Network in term of knee OA stages detection.

In conclusion, the discussion of OA classification studies using quadtree analysis is presented in this section. With the respect to the discussion of each study mentioned above, the main findings of the whole chapter were comprised into three sets of experiment:

1) In OA detection study the most appropriated support threshold value (σ), that drive the learning to get the best performance, was given by threshold value of 10 (σ=10). For the OA stages detection study, the threshold value of 10 can drive the

study to get the best performance of classification. Thus, it can be notified that the support threshold value of 10 can performance well for both knee OA and knee OA stages classification study.

2) The best performance of graph based study on region of interest (ROIs) to be considered that the most appropriated ROI for both studies. In OA detection study and knee OA stage classification study, presented that the most appropriate ROI performed by *Dataset* 3 that can drives both studies to get the best result was recorded. It can be concluded that the clearness on joint space image can produced the well performance of knee OA detection and knee OA stages classification study.

3) The most appropriated learning method between Naïve Bayes and Bayesian Network method. As the record of knee OA detection study presented that both Naïve Bayes and Bayesian Network are well performed for learning classification, while in knee OA stages classification study illustrated that Naïve Bayes was slightly better performance that Bayesian Network. It can be noted that both Naïve Bayes and Bayesian Network are well performance for this chapter study.

5.8 Summary

In brief, this chapter presented the graph based approach on knee OA detection and knee OA stages detection study. For OA detection study base on quadtree was presented in Section 4.2 and OA stages detection was presented in Section 4.3. Base on the reported of each section shown that: (i) support threshold value of 10 are well suitable to drive the best learning record of both studies, (ii) *Dataset* 3 sub-image can make a best learning classification for both study, and (iii) two well-known classifier: Naïve Bayes and Bayesian Network can produce the best record of learning classification for both studies.

CHAPTER 6

OSTEOARTHRITIS CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

6.1 Introduction

In This chapter the introduction of the applying of deep learning approach, typically Convolutional Neural Network (CNN) for knee-OA detection and knee-OA stage classification are described.

In OA classification study, the fundamental idea of deep learning approach is applied to OA classification is reported in this section. The fundamental idea is to construct the classifier from a collection of ROIs using deep learning algorithmsof this section is illustrated the nature of each Whole knee and knee joint space X-ray image, using deep learning algorithm. The sub-images of Dataset B (As mentioned in chapter 3) are used in this study. With the known class of each sub-image, the Convolutional Neural Network with transfer learning process had been analyse to classify for each sub-image dataset. In this process, AlexNet that is the pretrained model can build classifier which could be used to analyses OA condition.

More specifically, the deep learning approach of transfer learning technique is discussed. In addition the application of CNN to given ROIs or sub-image is also presented. To get the joint space more clear for the study, the Otsu's algorithm was applied to knee joint space sub-image. Thus, three sub-image datasets were applied for this study. Once each sub-image dataset has been obtained the next stage of the data preparation phase is to consider as the input layer which the input layer was processed in the hidden layer then produce as the output layer for the classification result. The CNN was applied in the feature learning of the transfer learning process to remove the feature selection process. The major idea of the study is to adopt the transfer learning technique, especially a Convolutional Neural Network. In case of medical image, in work (Tajbakhsh, et al., 2016) have been applied CNN for Carotid intima-media thickness (CIMT) test detection. In order to deal with the work with respect in this research, the CNN used the AlexNet pretrained model to learning the object feature in order to detect knee OA. A proposed framework of OA classification using CNN is presented in Figure 6.1bellow:

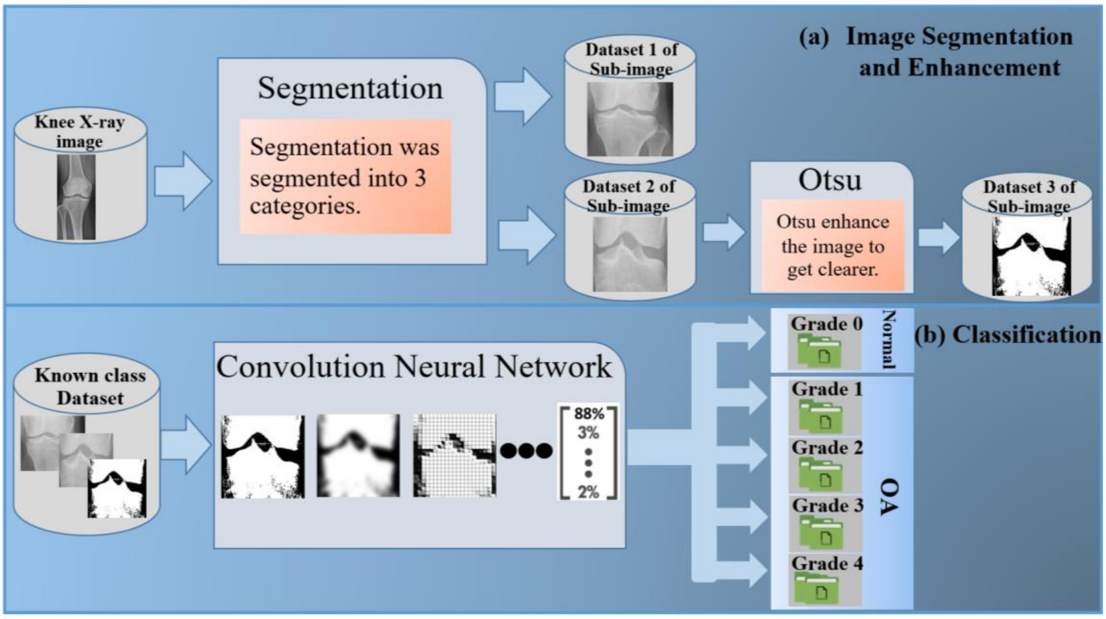


Figure 6.1 The CNN-Based of OA classification framework

Figure 6.1 illustrated that the transfer learning based of OA classification comprises of two main processes: (a) Image segmentation and Enhancement, and (b) Classification process. The image segmentation and enhancement, the Otsu thresholding method was considered. With respect to the work in this thesis, there are three sub-image datasets that obtained from the segmentation and enhancement process: (i) the whole knee sub-image considered as the *Dataset 1*, (ii) the joint space sub-image considered as the *Dataset 2*, and (iii) the enhancement of joint space sub-image (Otsu implementation) considered as the *Dataset 3*. Once a dataset of each sub-image has been segmented the next process is the transfer learning process with AlexNet pretrained model, the input image data needs to be resize to

227x227 pixels. An isolate a collection of image dataset used in learning process using the image feature in order to understand the nature of each image class.

The rest of this chapter is organised as follow: the information of convolutional neural network with transfer learning is presented in Section 6.2. The evaluation of the studies are demonstrated in Section 6.3 include Sub-section 6.3.1 presents the knee-OA detection evaluation and Sub-section 6.3.2 presents the evaluation of knee-OA stage classification study. The discussion of OA classification studies are presented in Section 6.4. Finally, the summary of the OA classification studies are presented in Section 6.5.

6.2 Convolutional Neural Network with AlexNet Transfer Learning

Deep learning is considered as one of the most famous algorithms of machine learning algorithms. Multi-levels learning have been applied. Deep learning is a technique developed from the artificial neural network which was inspired from human brain neurons connected system. Deep learning has been developed so far in computer vision technology including: image classification, object detection, and image segmentation. Deep learning model has had wiled development for various type of objection, the well-known deep learning model include (i) Autoencoder (AE), (ii) Deep Belief Learning (DBN), (iii) Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) (Du, et al., 2016). With the respect to the study purpose, the convolutional neural network is discussed.

Convolutional neural network (CNN or ConvNet) has been considered as the one of the most popular model for deep learning mechanism. In general CNN model is directly uses for identification and classification tasks of image, video, text, or sound. In the context of CNN application, the CNN directly used to find the patterns in image. Thus, CNN uses the pattern that multi levels learning can learn from image feature to classify image and in CNN work the feature extraction is removed from the process. There are three factors that make CNN widely used in deep learning for classification task: (i) the manual feature extraction is remove when applied with CNN, mean that CNN learn directly to the image feature, (ii) the state-of-art recognition result is produced by CNN, and (iii) when a new recognition task come, the CNN can be retrained and enabling to create on pre-existing network. CNN is similar to other deep learning model. It consists of three important layers: (i) Input layer, (ii) Hidden layer, and (iii) Output layer. The Figure 6.2 illustrates the three layers of CNN deep learning model:

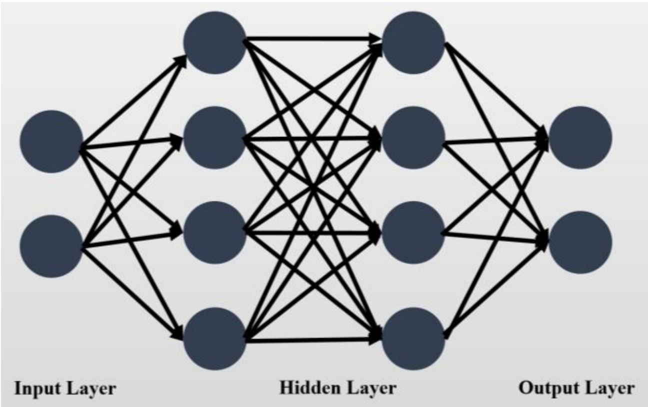


Figure 6.2 The Three Layers of CNN Deep Learning Model

With the reference to Figure 6.2, the three layers can be operated to the intent of learning feature. In the learning feature process, CNN use the most three common layers: (i) Convolution, (ii) activation or Rectified linear unit (ReLU), and (iii) Polling. Each of these three layer are described in further detail as follow:

- Convolution: the input is brought through a group of convolution filters that can activate certain image feature.
- ReLU: the maintaining positive value and mapping of negative value to zero is presented in order to produce the faster performance and more effective training. In this layer, only the activated features are selected and bring into the next layer that sometime called as activation.
- Pooling: the number of parameter is reduced that the network need to learn this can be call the output of nonlinear downsampling.

CNN application can be learnt by creating or building the CNN from scratch or use the pretrained model with study dataset. With the respect to this study, the pretrained model has been applied, typically AlexNet is the pretained model was used for the knee OA classification. AleNet is the CNN pretrained model that has been trained on 1.2 million high resolution image from the ImageNet LSVRC-2010 dataset (Alex, et al., 2012). The AlexNet model comprises of 23 layers and can be classified 1000 different categories (eg: head, knee, vehicle, cat, dog and etc.). AlexNet contained of 60 million parameters, five convolutional layers, in some of convolutional layer followed by max-pooling layers and three fully connected layers consists of 1000-way softmax that can separated 1000 different image groups (Alex, et al., 2012). AlexNet pretrained layers are presented in Table 6.1, layer 1 refers to the input image, and layer 25 is the classification layer. Thus, there are only 23 layers for AlaxNet pretrained network:

Table 6.1 The 23 Layers of Alex pretrained Network

Layer	Description
1 'conv1' Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0]
2 'relu1' ReLU	ReLU
3 'norm1' Cross Channel Normalization	cross channel normalization with 5 channels per element
4 'pool1' Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
5 'conv2' Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2]
6 'relu2' ReLU	ReLU
7 'norm2' Cross Channel Normalization	cross channel normalization with 5 channels per element
8 'pool2' Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
9 'conv3' Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1]
10 'relu3' ReLU	ReLU
11 'conv4' Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1]

Table 6.1 The 23 Layer of Alex Pretrained Network (cont.)

Layer	Description
12 'relu4' ReLU	ReLU
13 'conv5' Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1]
14 'relu5' ReLU	ReLU
15 'pool5' Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
16 'fc6' Fully Connected	4096 fully connected layer
17 'relu6' ReLU	ReLU
18 'drop6' Dropout	50% dropout
29 'fc7' Fully Connected	4096 fully connected layer
20 'relu7' ReLU	ReLU
21 'drop7' Dropout	50% dropout
22 ' ' Fully Connected	5 fully connected layer
23 'prob' Softmax	softmax

As noted in Chapter 1 Section 1.3 the OA classification study comprised of knee-OA detection and knee-OA stage classification. Thus, AlexNet have been applied with the 128 medical X-ray images OA classification studies. The CNN Alex pretrained of knee X-ray image process is illustrated in Figure 6.3

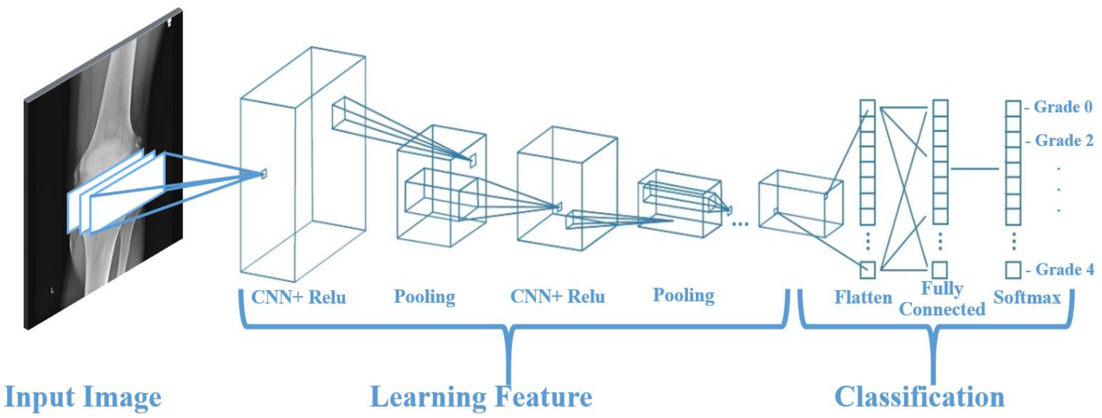


Figure 6.3 The example of CNN with many convolutional layers applied to Knee X-ray Imageries

Figure 6.3 illustrates that the input of images have to be 227x277 square image size, then CNN was applied to learning the feature identify the different image features. After the learning feature that has contain of most three common layers mentioned above, finished in many layers, the architecture of CNN go forward to classification step. In the classification step, the next-to-last layer of classification is the fully connected layer that produce the vector of K dimensions where K is number of classes that network can be used in the prediction of the probability value. The last layers of classification is defined by softmax, this final layer of CNN uses to provide the classification output.

Transfer learning is the commonly applying of deep learning application. In this study, the transfer learning of knee OA classification applied with AlexNet is discussed. In transfer learning, the pretrained network can take whenever the new classification task begin. Transfer learning work quickly and better performance of small image dataset. The Figure 6.4 illustrates the transfer learning mechanism of the study.

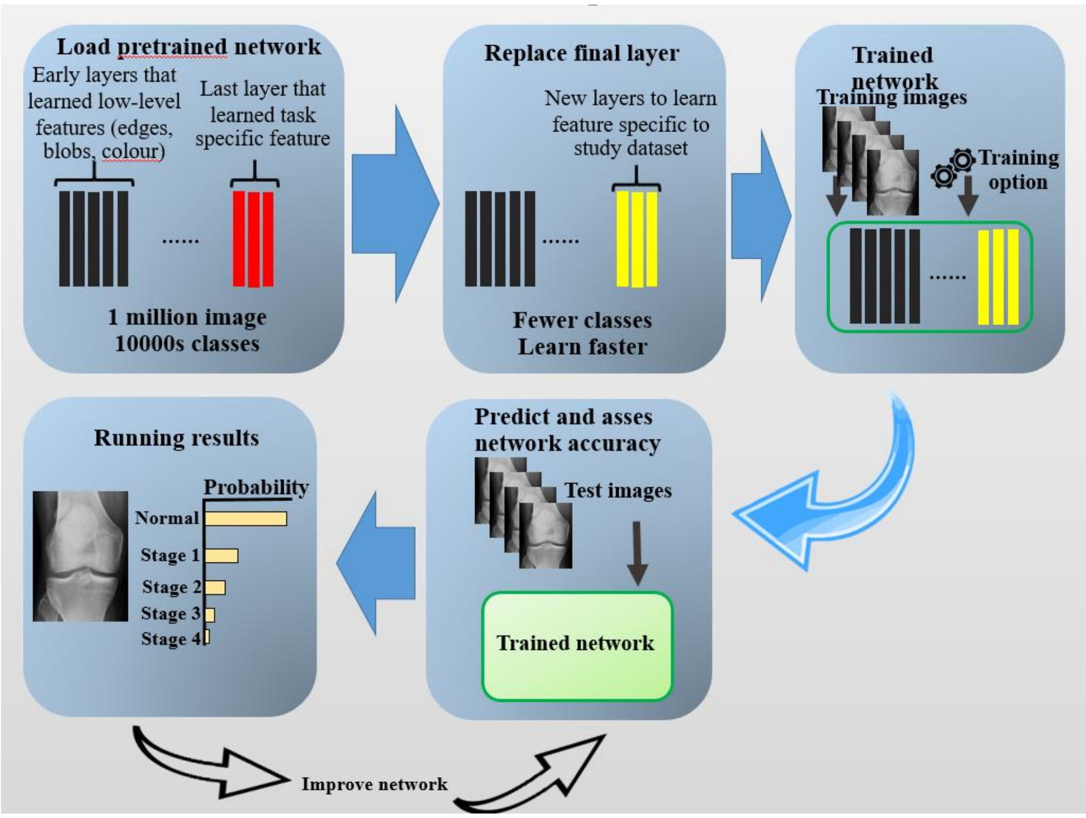


Figure 6.4 The example of transfer learning to four classes of medical image

From figure 6.4, it can be seen that the process of transfer learning with AlexNet pretrained model. The first process is the applied of AlexNet which has 1 million images to apply the new task of medical image classification, then the second process is the new layer which used to learn to specific to image dataset. Next, the trained network process which is the process taking long time generate of the transfer learning process which can be learnt image feature for identifying the image class, then the predict and asses network accuracy has been applied to each image as the probability value of classification result. Finally, the unseen image could be classified as the probability value. The evaluation of knee-OA detection and knee-OA stage classification using the transfer learning is presented in Section 6.3 as follow.

6.3 Evaluation

As noted before the research study was divided into two main objectives: (i) OA detection and (ii) OA stage Classification. The detail of evaluation is described in Sub-section 6.3.1 and 6.3.2 respectively.

6.3.1 Osteoarthritis Detection using CNN

For knee-OA detection using CNN is presented here, the evaluation of the transfer learning with AlexNet to detect knee OA is presented in this sub-section. Once, with the three different sub-image dataset have been applied to the study include: (i) Set 1 sub-image (the image of whole knee segmented image), (ii) Set 2 sub-image (the knee joint space segmented image), and (iii) Set 3 sub-image (the implementation of Otsu's method to Set 2 sub-image). As a result, the result of transfer learning all the sub-image dataset (128 sub-images for each dataset) to knee OA detection study with the value of accuracy of 0.88 (Set 1: AC =0.88, Set 2: AC= 0.88, Set 3: AC= 0.88). However, the study have been extended to study of CNN on the random image of the dataset of (i) 8 images, (ii) 16 images, (ii) 32 images and (iv) 64 images. Table 6.2 reports the accuracy of each dataset:

Table 6.2 The Accuracy value of CNN application for OA detection

No. of image	8	16	32	64	128
Dataset 1	0.5	0.75	0.8333	0.8333	0.88
Dataset 2	1	1	1	1	0.88
Dataset 3	0.5	0.5	1	0.75	0.88

From Table 6.2 it can be seen that the dataset 2 sub-image work well for CNN approach with different image dataset (8 image, 16 image and 64 images produce the best accuracy with full value of 1). From the result in Table 6.2, the three datasets result can be plotted in the graph which illustrated in Figure 6.5:

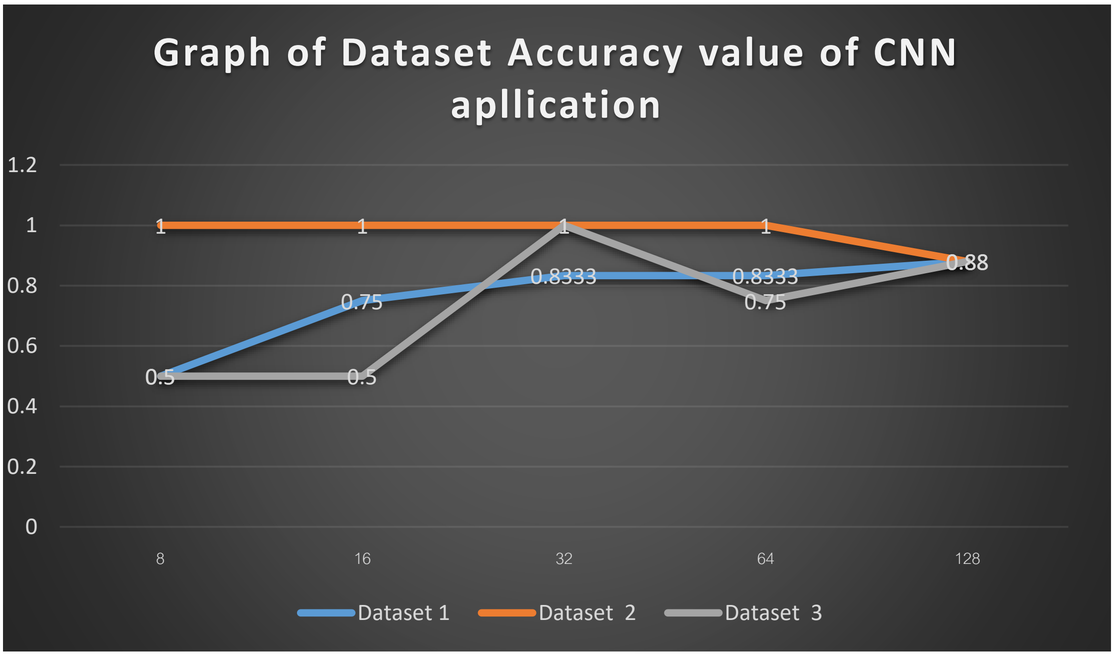


Figure 6.5 The graph of accuracy for each dataset of CNN application for OA detection

6.3.2 Osteoarthritis Stage Classification using CNN

The evaluation of the transfer learning with AlexNet model to knee-OA stage classification study is presented in this sub-section. With the respect to the three different sub-image dataset have been applied to the study include: (i) *Dataset* 1 sub-image (the image of whole knee segmented image), (ii) *Dataset* 2 sub-image (the knee joint space segmented image), and (iii) *Dataset* 3 sub-image (the implementation of Otsu method to algorithm 2 sub-image). As a result, the result of transfer learning all the algorithm sub-image to knee OA detection illustrated in Table 6.3 below:

Table 6.3 The Accuracy value of CNN application for OA stage classification

No. of image	16	32	64	128
Dataset 1	0.4	0.2	0.5	0.5185
Dataset 2	0.4	0.4	0.3571	0.4444
Dataset 3	0.4	0.2	0.5	0.6296

From Table 6.3 it can be seen that the dataset 3 sub-image work well for CNN approach with 128 image dataset. From the result in Table 6.3, the three datasets result can be plotted in the graph which illustrated in Figure 6.6:

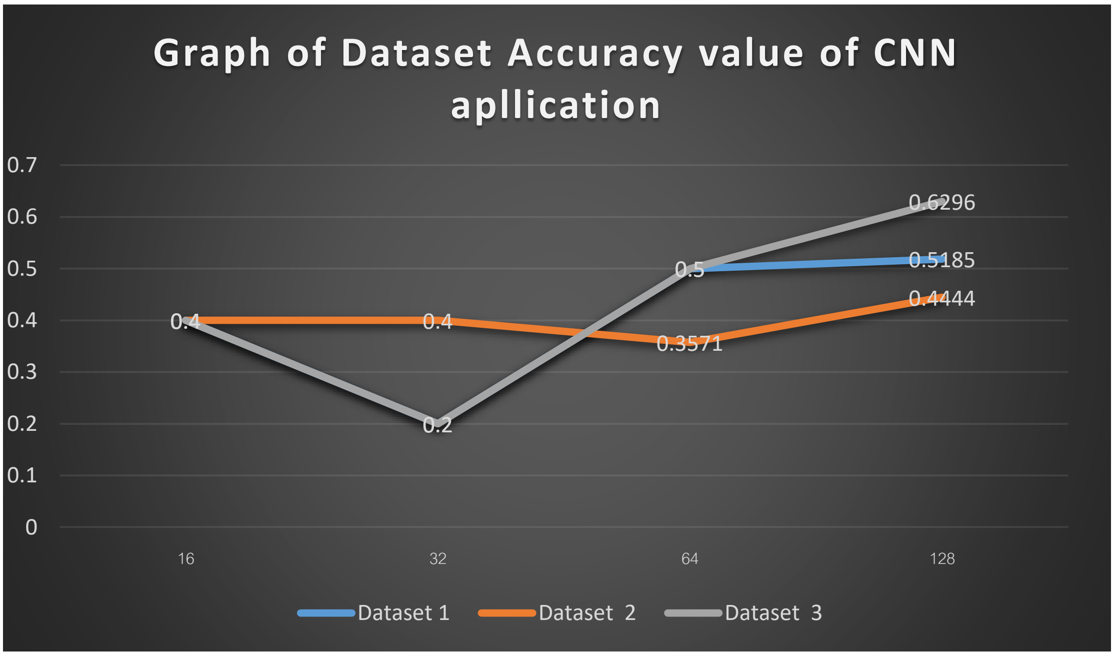


Figure 6.6 The graph of accuracy for each dataset of CNN application for OA Stage Classification

From Figure 6.6 it can be Observed that the graph of Dataset 3 produce the best result when the image is getting higher and higher. The Dataset of each Dataset is separated a number of image include 16 images, 32 images, 64 images and 128 images. The Table 6.4 reports the number of each OA stage for CNN application of each image dataset.

Table 6.4 The number of OA stage to each image dataset of 16, 32, 64 and 128 images

Stage OA / Number of image	**0**	**1**	**2**	**3**	**4**
16	3	3	4	3	3
32	7	7	6	6	6
64	14	14	14	14	8
128	38	17	21	44	8

6.4 Discussion

In this section the discussion of knee-OA classification is presented, the CNN deep learning mechanism with AlexNet pretrained model of transfer learning on OA detection and OA stage classification were considered in this chapter. With the respect to the discussion of Sub-section 6.3.1 and Sub-section 6.3.2, the main findings of chapter presented that in the context of OA detection the *Dataset* 2 sub-image are well predicted for the study, while knee OA stages classification study are well performance by *Dataset* 3 sub-image (the clear joint space sub-image which applied Otsu's method), followed by the *Dataset* 1 sub-image (the whole knee segmented image).

6.5 Summary

In short, the chapter presented the CNN deep learning approach with AlexNet transfer learning on knee OA detection and knee OA stages detection study. Thus, in the chapter mainly focused two study approach: (i) OA detection and (ii) OA stages detection. For OA detection study base on transfer learning with AlexNet was presented in Sub-section 6.3.1 and OA stages detection was pictured in Sub-section 6.3.2. Base on the reported of each section shown that: for *Dataset* 2 sub-image were well produced for knee OA detection, while the *Dataset* 3 sub-image dataset was well produce for knee OA stages classification study.

CHAPTER 7

CONCLUSION

The summary of the proposed knee OA detection study and knee OA stages classification study using medical X-ray imagery, the main findings, the research contributions and direction for possible future study are presented in this chapter. The chapter is organised as follow: the summary of the proposed knee OA and Knee OA stage classification presented in this thesis discussed in Section 7.1. While Section 7.2 presents the main findings and research contributions of the research work. Finally, Section 7.3 discusses some guideline for the future research study.

7.1 Summary

In this thesis, there are three different approaches (texture-based, graph-based, and CNN) were proposed to classify knee OA and Knee OA stages with the respect to the ROIs of knee X-ray images. For the evaluation of these three approaches, the first approach applied four sub-images, when three sub-images datasets were applied for the second and the third approach. In each case an ROI identification was applied in the knee region of the X-ray image. With the respect to the original medical X-ray image presented in DICOM format, it produce a large size, thus, the segmented ROI produced the benefit of small size and unique for study to each approach. For each approach was finished in a different field: (i) texture-based, (ii) graph-based and (iii) convolutional neural network based (CNN-based).

The first approach in this thesis was founded by the texture based approach. In this approach the first step was considered as the ROI segmentation. With the respect to the ROI segmented process was illustrated, there are four ROIs were identified from each knee X-ray image. Once, the ROIs were obtained is then applied. In the context of the work presented in this research, there were ten texture descriptors were considered. When the feature vector was completed, the feature selection (the process of feature space reduction) was then applied in order to build a feature space that suitable to classifier generation process. In the feature selection process, there are five well-known methods were used in the study, while in classification process, nine learning classifier learning algorithms were introduced. The evaluations: (i) in knee-OA detection, the best recorded classification accuracy was obtained from the femur bone region with neural network algorithm with CFS feature selection method; and (ii) in knee OA stage classification, the best result was obtained by the Tibia bone region with the logistic regression learning together with CFS feature selection.

The second approach was founded by the application of graph based to medical knee X-ray imagery. More specifically, the ROIs were identified. Once each ROI was segmented, the individual ROI was decomposed using quadtree. However, the quadtree representation does not depend on itself to incorporation with reference to learning method. In this case, the subgraph mining was applied to identify frequently occurring subgraph which can be used as feature in term of feature vector representation. The identified frequent subgraphs were view as defining a feature space which could be used to represent the image dataset. A given image dataset could thus be recast into this format so that each image is represented by a feature vector whose elements were some subset of the global set of identified frequent subgraph making up a feature space. The feature vector can be used directly to the learning classifier for the classification. The reported evaluation indicated that both knee OA detection and knee OA stages classification. The best accuracy results were obtained when applying low support threshold value for identifying occurring subgraph.

The third approach was founded by the concept of deep learning approach to medical image classification by using of convolutional neural network in transfer learning of AlexNet pre-trained model. More specifically, the ROI were segmented into three ROIs, where each ROI need to be resize to *227x227* that the learning need to be trained for AlexNet. In the application of CNN model, the feature extraction process was removed which instead by learning feature by the layer of AlexNet model. When the model learned each ROI image feature, the classification will be classify by the last layer of AlexNet model. The reported evaluation illustrated that for knee OA detection, the best classification accuracy was obtained by all the ROIs were segmented, while in knee OA stages classification best result was obtained from *Dataset* 3 (Otsu application to knee joint space) with neural network.

7.2 The Main Findings and Research Contributions

In this section, the main findings of the thesis and some contributions are presented. The research presented in this thesis was provided directly with an answer to the research question presented in Chapter 1, namely:

Can the knee OA and the stage knee OA can be predicted by applying classification technique and deep learning model to human joint X-ray imagery?

This research question had a number of research issues that need to be resolution before an answer to the central research question could be derived. With respect to this thesis work, the main findings are presented with associated to research question and issues. This section is arranged by considering each of the research issues itemised in Chapter 1 in turn and then returning to the research question.

1) What is the most appropriate ROI for classification study in case of knee OA detection study and knee OA stages detection study?

Three were two groups of ROI segmentation in this thesis. For the first ROI group was divided into four ROIs mentioned in Chapter 3 which was used in texture-based approach. For the second ROI group (SET B) is consisted of three ROIs mentioned in Chapter 3 which was used in graph-based approach and deep learning approach. The first was used in Chapter 4 and comprised three main process for the study (For knee OA detection the femur region performed well, while in knee OA stages the Tibia region produced the most accuracy for classification result). For the second

ROI group was applied in Chapter 5 graph-based approach and Chapter 6 deep learning approach. In graph based approach presented in Chapter 5, ROI was used with quadtree decomposition. In deep learning model study, ROI was used with CNN of transfer learning of AlexNet pretrained model. With the respect to the reported evaluation of best performance of both knee OA and knee OA stages detection in both graph-based and deep learning were performed by the Dataset 3 3 sub-image (the ROI segmented of knee joint space applied with Otsu method to make a clear joint space).

2) What is the most appropriate feature extraction method in case of texture based approach for both knee OA and knee OA stages detection study?

In this finding is presented only in texture based approach which was demonstrated in Chapter 4. Obtained ROI was presented the texture by texture descriptors. Ten texture descriptors were suggested to the work presented in this work in order to generate feature vector which was later used in classification. Amounts of ten texture descriptor the Local Binary Pattern (LBP) and its' family technique was the most effective of the study. As the study reported from the evaluation in chapter 4, LBP was the best texture descriptor technique for both knee OA and Knee OA stages classification applications.

3) What is the most appropriate feature selection techniques in case of texture based approach for knee OA detection study?

With the respect to Chapter 4 of texture based approach, after the texture descriptors were applied, the outcome of the descriptor was the feature space which contained of a number of features. In this case, feature selection were applied to reduce feature space dimensionality to make feature vectors that suitable to use directly with learning methods. The five well-known feature selections were applied in Chapter 4. As the reported of the evaluation illustrated that Correlation-based Feature Selection (CFS) was the best technique in context of knee OA study. Thus, CFS was applied in knee OA stages classification study and Chapter 5 about the graph-based approach and not further detail more of other feature selection techniques due to the CFS performance was the highest compare to others.

4) What is most appropriate classification techniques for predicting knee OA and knee OA stages?

The nine classifier generation methods were apply in texture based approach was presented in Chapter 4. When the feature selection process was completed, the reduction feature vector and the classifier generation was then commenced. With the study reported that in case of knee OA detection study, the best three of classifier generation mechanisms were Bayesian Network, then followed by Logistic Regression and the last was presented by Naïve Bayes classifier. For knee OA stages classification, the best three method illustrated that the best method was presented by Logistic Regression, followed up by Bayesian Network then finished by Naïve Bayes classifier.

5) In case of graph-based approach, what is most appropriate support threshold value for predicting knee OA and knee OA stages study?

There were five support threshold values presented in Chapter 5 of graph-based approach. With the application of quadtree decomposition and subgraph mining to generate the feature vector with CFS feature selection. The support threshold is the primary factor that is important for frequent subgraph mining. As the reported of the evaluation, the low support threshold values was driven the best performance to both study of knee OA and knee OA stages detection.

6) Is the performance of deep learning model powerful for predicting knee OA and knee OA stages study?

In the deep learning model to medical image analysis has not been widely used. Thus, the efficiency of deep learning for knee OA and knee OA stages classification was also considered as the main finding in this thesis work. With the application of CNN deep learning using AlexNet pre-trained model, the feature extraction process was removed. Therefore, the hidden layer of CNN worked as the learning feature and classification. The result of the study presented in Chapter 6 reported that, for knee OA detection, CNN was powerful to deal with the task, while in the knee OA stages detection, the predicted accuracy was accepted with the application of CNN.

Back to the initial research question, the knee OA and the stage of knee OA can be predicted by applying classification technique and deep learning method to human joint X-ray imagery can be founded on the process that encompasses: (i) in case of texture-based study, for knee OA detection, the Femur bone region (ROI) performed a well prediction, while in knee OA stages classification the Tibia bone region (ROI) can produced a well prediction. In case of graph-based approach and deep learning approach the Otsu knee joint space segmented image was suggested, (ii) in case of texture-based approach, the LBP feature descriptor was preferred for knee OA and knee OA stage classification, (iii) CFS was the most appropriate feature selection that produced the best classification results of knee OA and knee OA stages, (iv) in the context of learning classification techniques Bayesian Network was the most appropriate technique for OA detection technique, while Logistic Regression performed the best result for OA stage classification, (v) in graph-based approach, the low support threshold produced the best classification performance, (vi) deep learning approach performed a good classification performance on OA detection, however it was not well with OA stage classification (highest AUC of 0.6296). The experiment results indicated that a good prediction of knee OA and knee OA stages could be obtained at very little cost.

The primary contribution of the research work illustrated in this thesis where pictured in Section 1.4 of Chapter 1, for convenient all the contributions are again discussed below. Note that in each case of the related chapter where the contribution was establish is given in parenthesis.

1) A knee sub-image (ROI) representation founded on the concept of “texture” analysis. More specifically applying of Local Binary Patterns (LBPs), as before a feature vector format was build.

2) A knee sub-image representation founded on the concept of “graph based” by applying the quadtree hierarchical decomposition together with frequent subgraph mining for reducing the feature dimensionality. The identifier frequent subgraph were set to a feature vector format, one vector per ROI, suited for input into a learning classifier.

3) An approach of deep learning approach was done without manually feature extraction process.

4) An analysis of a sequence of the proposed sub-image (ROI) so as to identify the most appropriate ROIs in term of knee OA classification from X-ray images.

5) An analysis of a sequence of the proposed ROI feature extraction algorithm so as to select the most appropriate in term of knee OA classification from X-ray images.

6) An analysis of a sequence of feature selection algorithm so as to select the most appropriate in term of knee OA classification from X-ray images.

7) An analysis of a sequence of classifier generation algorithm so as to select the most appropriate in term of knee OA classification from X-ray images.

7.3 Future Works

The research pictured in this thesis has a number of the guideline that can be considered for the future works. In the concluding of this thesis, and this chapter, theses future works guideline are summarised as below:

1) Perceptual Browsing Component (PBC) and Similarity Retrieval Component (SRC) texture Descriptor for feature extraction.

In term of the texture based approach presented in Chapter 4, the texture feature of each ROI was extracted by 10 texture descriptors (LBP, LBP-HF, RLBP, CLBP, and etc,.). However, the texture feature can be extracted further more on the structuredness of feature by applying Perceptual Browsing Component (PBC) and Similarity Retrieval Component (SRC). A proposed study of addressing Perceptual Browsing Component (PBC) and Similarity Retrieval Component (SRC) for image texture analysis presented in work (Wu, et al.,1999).

2) The Weighted frequent subgraph mining

In the work of graph-based application illustrated in Chapter 5, frequent subgraph mining was used to identify frequently occurring subgraph which were used to build the feature vectors. However, the subgraph mining process works by allocating a count of 1 to present the region in knee image and 0 is outside the knee region; it take no account of the number of time it appears. The algorithm to address this is to adopt what is called as weighted subgrap mining. The work (Jiang, et al., 2010) presented the application of weighted frequent subgraph mining.

In conclusion the work illustrated in this thesis has demonstrated that it is possible to predict knee OA and knee OA stages within a given knee medical x-ray imagery at a much reduce cost that that which would take with medical doctor to analyse.

3) Deep Recurrent Neural Networks for knee Osteoarthritis detection and classification.

With the application of CNN deep learning presented in Chapter 6, CNN performed well for Knee OA detection, and acceptable accuracy for knee OA stages classification. Thus, Deep Recurrent Neural Networks (RNN) which one of the famous deep learning model may be good direction for the study in this thesis. A mechanism of RNN have been presented in (Mou, et al., 2017, Maggiori, et al., 2017 and Wang, et al., 2016) for image analysis.

In conclusion the work illustrated in this thesis has demonstrated that it is possible to classify knee OA and knee OA stages within a given knee medical x-ray imagery at a much reduce cost that that which would take with medical doctor to analyse.

References

Agrawal, R., and Srikant, R., "Fast Algorithm for mining association rules in large database", Proceeding of The Twentieth International Conference on Very Large Data Bases (VLDB), pp. 487-499, San Francisco, CA, USA, 1994, Morgan Kaufmann Publishers Inc.

Alex, K., Ilya S., and Geoffrey E. H., "ImageNet Classification with Depp Convolutional Neural Networks", Proceeding of The 25th International Conference on Neural Information Processing Systems, vol 1, pp. 1097-1105, 2012.

Alejandro, C. F., and Perciano, T., "Introduction to Image Processing Using R", 2013, pp. 21-30

Anitha, N., Keerthika, G., Maheswari, M., and Praveena, J.,"A framework for medical images classification using soft set", Proceeding of the Second International Conference on Current Trends in Engineering and Technology (ICCTET), pp. 268-272, July 2014.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., and Mougiakakou, S., "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1207-1216, 2016.

Bandyopadhyay, S.K., "Edge Detection in Brain Images", Proceeding of The (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), pp. 884-887, 2011.

Bindu, Ch. H., "An Improved Medical Image Segmentation Algorithm Using Otsu Method", International Journal of Recent Trends in Engineering, Vol 2, No. 3, pp. 88-90 November 2009.

Burl, M. C., Fowlkes, C., and Roden, J., “Mining for image content,” Systemics, cybernetics, and informatics/information systems: analysis and synthesis, 1999.

Chen, Q., and Koltun, V., “A Simple Model for Intrinsic Image Decomposition with Depth Cues,” Proceeding of The IEEE International Conference on Computer Vision, Sydney, NSW, 2013, pp. 241-248.

Cho, H., Chang, C., Kim, K., Park, J., Yoo, J., Koh, I., and Kim, T., “Gender and prevalence of knee osteoarthritis types in elderly Koreans” J Arthroplasty. 2011; 26(7): 994–999

Cole, J. H, Poudel, R.PK, Tsagkrasoulis, D., Caan, M. WA, Steves, C., Spector, T. D., and Montana, G., “Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker,” NeuroImage, vol 163, pp. 115-124, 2017

Cortes, C., and Vapnik, V., “Support-Vector Networks”, Machine Learning 20 (3) (1995) 273–297.

Dryden, I. L., and Mardia, K. V., “Statistical Shape Analysis. John Wiley & Sons”, 1998.

Dua, S., Kandiraju, N., and Chowriappa, P., “Region Quad-Tree Decomposition Based Edge Detection for Medical Images.” The Open Medical Informatics Journal 4 (2010): 50–57. PMC. Web. 10 Sept. 2018.

Du, X., Cai, Y., Wang, S., and Zhang, L., "Overview of deep learning," Proceeding of The 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 159-164.

Esteva, A., Kuprel, B., Novoa, R., Ko, J., Swetter S., Blau, H., and Thrun, S., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, Macmillan Publishers Limited, vol. 542, no. 7639, pp. 115-118, 2017.

Fawcett, T., “An introduction to ROC analysis”. Science direct on Pattern Recognition Letter, vol 27, pp.861–874, 2006.

Fayyad, U., Shapiro, G., Smyth, P., and Uthurusamy R., Advances in Knowledge Discovery and Data Mining 1996, American Association for Artificial Intelligence.

Fesharaki, N., and Pourghassem, H., “Medical X-ray image classification based on shapes features and Bayesian rule.”, Proceeding of the Fourth International Conference on Computational Intelligence and Communication Networks (CICN), pp. 369-373, November 2012.

Firoz, R., Ali, Md. S., Khan, M.N.U., Hossain, Md.K., Islam, Md.K., and Shahinuzzaman Md., “Medical Image Enhancement Using Morphological Transformation.”, Journal of Data Analysis and Information Processing, pp. 1-12, Feb. 2016.

Fu Yongjian, “Data Mining: Task, Techniques, and Application”, Cleveland State University, retrieved from “http://academic.csuohio.edu/fuy/Pub/pot97.pdf”, accessed December 26th 2016

Fransen, M., Bridgett, L., March, L., Hoy, D., Penserga, E., and Brook, P., “The epidemiology of osteoarthritis in Asia”, Int J Rheum Dis. 2011; 14: 113-121.

Friedman, N., Murphy, K., and Russell, S., “Learning the structure of dynamic probabilistic networks”, Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, pp. 139-147, July 1998

Gonzalez, R. C., Woods, R. E., and Eddins, S. L., “Digital Image Processing Using Matlab”. (IND,2004).

Gornale, S., Patravali P., and Manza, R., "Detection of Osteoarthritis using Knee X-Ray Image Analyses: A Machine Vision based Approach", International Journal of Computer Application, Vol 145-No.1, pp. 20-26, July 2015.

Guo, Z., Zhang, L., and Zhang, D., "A completed modeling of local binary pattern operator for texture classification", IEEE Transactions on Image Processing 19 (6) (2010) 1657–1663.

Guo, Y., Zhao, G., and Pietik¨ainen, M., "Texture classification using a linear configuration model based descriptor", Proceedings of the British Machine Vision Conference (2011) 199.1–199.10.

Haddad, M., and Kheddouci, H., "Graph based approaches for service oriented applications in ad hoc networks", Proceeding of The IEEE international Conference on Pervasive Services (ICPS), pp. 431-436, July 2007.

Hall, M., "Correlation-based Feature Selection for Machine Learning", Methodology (April) (1999) 1–5.

Han, J., and Kamber, M., Data Mining : Concepts and Techniques 2006, Morgan Kaufman Publishers, USA.

Han, J., and Kamber, M., "Data mining: concepts and techniques", Morgan Kaufmann Publisher Inc., San Francisco, CA, USA, Third Edition, 2011.

Han, J., Kamber, M., and Pei, J., Data Mining: Concepts and Techniques. 2012.

Havaei, M., Davy, A., Farley, D. W., Biard, A., Courville, A., Bengio, Y., Pal, C., and Jodoin, P.M., "Brain tumor segmentation with deep neural networks," Medical image analysis, vol. 35, pp. 18-31, 2017.

Herman, A., Chechik, O. , Segal, G., Kosashvili, Y., Lador, R. , Salai, M., Mor, A., Elbaz, A., and Haim, A., "The correlation between radiographic knee OA and clinical symptoms-do we know everything?", Clin Rheumatol (2015), vol 34, pp.1955-1960, 2015.

Huan, J., Wang, W., and Prins, J., “Efficient mining of frequent subgraphs in the presence of isomorphism”, Proceeding of The Third IEEE International Conference on Data Mining (ICDM), pp. 549-552, Florida, USA, Nov 2003.

Inokuchi, A., Washio, T., and Motoda, H., “An Apriori-based algorithm for mining frequent substructures from graph data,” Proceeding of The International Conference on Principle and Practise of Knowledge Discovery in Databases (PKDD), pp. 13-23,2000.

Jiang, C., FCoenen, F., and Zito, M., “Frequent sub-graph mining on edge weighted graphs,” Proceeding of The Twelfth International Conference on Data Mining and Knowledge Discovery (DaWak). pp. 77-88, Bilbao, Spain, Aug 2010.

Jin, C., Yang, Y., Xue, Z. J., Liu, K.M., and Liu, J., “Automated Analysis Method for Screening Knee Osteoarthritis using Medical Infrared Thermography”, Journal of Medical and Biological Engineering 33, vol 5, pp. 471-477, Feb.2012.

Kellgren, J., and Lawrence, J., “Radiologic assessment of osteoarthritis”, Annals of the Rheumatic Diseases, vol. 16, pp. 494–501, Dec. 1957.

Kellgren, J., Jeffrey, M., and Ball, J., “Atlas of Standard Radiographs”, Blackwell Scientific, Oxford, UK, 1963.

Khalili, F., Celenk, M., and Akinlar, M.A., “Medical Image Compression Using Quad-tree Fractals and Segmentation.” 2013

Kira, K., and Rendell, L., “The feature selection problem: Traditional methods and a new algorithm”, Aaai journal, 1992, pp.129 – 134

Kira, K., and Rendell, L.,   A practical approach to feature selection. (1994).

Kooi, T., Litjens, G., Ginneken, B.V., Mérida, A. G., Sánchez, C. I., Mann, R., Heeten, A.D., and Karssemeijer, N., “Large scale deep learning for computer aided detection of mammographic lesions,” Medical image analysis, vol. 35, pp. 303-312, 2017.

Kubakaddi, S., Ravikumar, K., and Harini DG, “Measurement of Cartilage Thickness for Early Detection of Knee Osteoarthritis (KOA)”, 2013 IEEE Point-of-Care Healthcare Technologies (PHT), Bangalore, pp. 208-211, Jan. 2013.

Kuramochi, M., and Karypis, G., “Frequent subgraph discovery”, Proceeding of The IEEE International Conference on Data Mining (ICDM), pp. 313-320, 2001.

Lai, Z., Qu, X., Lui, Y., Guo, D., Ye, J., Zhan, Z. and Chen. Z., “Image reconstruction of compressed sensing MRI using graph based redundant wavelet transform”, Medical Image Analysis, 2015.

LeC, Y., Haffner, P., Bottou, L., and Bengio, Y., “Object Recognition with Gradient-Based Learning”.1999.

Lei, S., “A Feature Selection Method Based on Information Gain and Genetic Algorithm”, Proceeding of The International Conference on Computer Science and Electronics Engineering (2012) 355–358.

Lewandowski, C. M., “The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference”, (2015) 1689–1699.

Lisle, D., “Imaging for Students Fourth Edition”, (UK, 2012).

Maggiori, E., Charpiat, G., Tarabalka, Y., and Alliez, P., “Recurrent Neural Networks to Correct Satellite Image Classification Maps,” Proceeding of The IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 9, pp. 4962-4971, Sept. 2017.

Maintz, T., “Digital and Medical Image Processing”, Inleiding Beeldverwerking Journal, 2005, (10): 18-24.

Mehta, R., and Egiazarian, K., “Rotated Local Binary Pattern (RLBP) Rotation Invariant Texture Descriptor”, Icpram (2013) 497–502.

Mehta, R., and Egiazarian, K., “Dominant Rotated Local Binary Patterns (DRLBP) for texture classification”, Pattern Recognition Letters 71 (2016) 16–22.

Mittal, N., Singh, H.P., and Gupta, R., “Decomposition & reconstruction of medical images in MATLAB using different Wavelet parameters,” Proceeding of The International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Noida, 2015, pp. 647-653.

Moeskops, P., Viergever, M. A., Mendrik, A. M., de Vries, L. S., Benders, M. J. N. L., and Išgum, I., “Automatic segmentation of MR brain images with a convolutional neural network,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1252-1261, 2016.

Mou, L., Ghamisi, P., and Zhu, X. X., "Deep Recurrent Neural Networks for Hyperspectral Image Classification," Proceeding of The IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 7, pp. 3639-3655, July 2017.

Ng, H., Huang, S., Ong, S., Foong, K., Goh, P.S., and Nowinski, W.L. , “Medical Image Segmentation Using Watershed Segmentation with Texture-Based Region Merging”, Proceeding of The 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, pp. 4039-4042, Oct. 2008.

Nielsen, and Michael, A., (2015). Neural Networks and Deep Learning, Chapter 6, 2015.

Ojala, T., Pietik¨ainen, M., and Harwood, D., “A comparative study of texture measures with classification based on featured distributions”, Pattern Recognition 29 (1) (1996) 51–59.

Olmos, I., and Gonzalez, J.A., “Structure graph-based representations used for finding hidden pattern”, Fourth Latin America Workshop on Non-Monotonic Reasoning (LANMR), Puebla, Mexico, Oct. 2008.

Pandey, M., Soam, S., and Tripathi, S., “Detection of Knee Osteoarthritis Using X-Ray”, Proceeding of ACEIT Conference, pp 1-4, 2016.

Pereira, S., Pinto, A., Alves, V., and Silva, C. A., “Brain tumor segmentation using convolutional neural networks in MRI images,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1240-1251, 2016.

Plackett, R. L., and Pearson, K., The Chi-squared Test, 51 (1983) 59–72.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., and Nielsen, M. , "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," Proceeding of The International conference on medical image computing and computer-assisted intervention, 2013, pp. 246-253.

Quinlan, J. R., “Induction of Decision Trees.”, Machine Learning 1 (1) (1986) 81–106.

Riaz, F., and Ali, K.M., “Application of Graph Theory in Computer Science”, Proceeding of The Third International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), pp. 142-145, July 2011.

Roweis, S. T. and Saul, L. K, "Nonlinear Dimensionality Reduction by Locally Linear Embedding". Science. Vol 290 (5500), pp. 2323–2326. Dec. 2000.

Ryan, J. U., Melissa, M., William, L., Randal, S. O., Jason, H. M., “Relief-Based Feature Selection: Introduction and Review”. Journal of Biomedical Informatics. (2017) 85.

Sahu, A., Bhateja, V., Krishn, A., and Himanshi, “Medical Image Fusion with Laplacian Pyramids”, Proceeding of The International Conference on Medical Imaging, m-Health and Emerging Communication Systems. Pp. 448-453. 2014

Samet, H., “The quadtree and related hierarchical data structure”, Journal of ACM Computing Survey (CSUR). 16(2). pp. 187-260. June 1984.

Samet, H., “Hierarchical spatial data structure,” Proceeding of The proceeding of the First Symposium on Design and Implementation of Large Spatial Databases, pp. 1993-212.UK, 1990.

Sargunar, P., and Sukanesh, R., “Texture Defect Detection Technique for MRI Knee Images”, ACEEE International Journal on Signal and Image Processing, Vol 1, No. 2, pp. 48-52, Feb. 2009.

Senthilkumaran, N., and Vaithegi, S., “Image Segmentation By Using Thresholding Techniques For Medical Images”. Computer Science & Engineering: An International Journal (CSEIJ), Vol.6, No.1, pp. 1-13, February 2016

Shamir, L., Ling, S., Scott, W., Jr., Bos, A., Orlov, N., Macura, T., Eckly, D., Ferrucci, L., and Goldberg, I., “Knee X-Ray Image Analysis Method for Automated Detection of Osteoarthritis”, in IEEE Transactions on Biomedical Engineering, vol. 56, no. 2, pp. 407-415, Feb. 2009.

Shamir, L., Ling, S., Scott, W., Hochberg, M., Ferrucci, L., and Goldberg I., “Early detection of radiographic knee osteoarthritis using computer-aided analysis”, Osteoarthritis and Cartilage (2009), Vol 17, pp. 1307-1312, Oct 2009.

Shamir, L., Rahimi, S., Orlov, N., Ferrucci, L., and Goldberg, I., “Progression Analysis and Stage Discovery in Continuous Physiological Process Using Image Computing”, EURASIP Journal on Bioinformatics and Systems Biology, Vol 2010, Art ID 107036, pp. 1-7, 2010.

Shamir, L., “A computer analysis method for correcting knee X-rays with continuous indicator”, International Journal of Computer Assisted Radiology and Surgery (2011), vol 6, pp. 699-704, March 2011.

Spacetelescope. (2016) “Introduction to Image Processing”, retrieved from “https://www.spacetelescope.org/static/projects/fits_liberator/image_processing.pdf”, accessed December 26th 2016.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, T.R., Kendall, B.C., Gotway, B.M., Liang, J., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299-1312, May 2016.

Wagaj, B. L., and Patil, M.M., “Osteoarthritis Disease Detection with the Help of Image Processing Technique”, Proceeding of The Emerging Application of Electronic System, Signal Processing and Computing Technologies Conference (NCESC 2015), pp 1-4, 2015.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., and Xu, W., “CNN-RNN: A Unified Framework for Multi-label Image Classification,” Proceeding of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2285-2294.

Wang, Y., Bahrami, S., and Zhu, S.C., “Perceptual scale space and its Application,” Proceeding of The Tenth IEEE International Conference on Computer Vision (ICCV), pp. 58-65. Beijing, China, Oct. 2015.

Waters, H., “The First X-ray,1895”, The Scientist, retrieve from “http://www.thescientist.com”, accessed on October 15 2018.

Watt, E. W., and Bui, A. A., “Evaluation of a dynamic bayesian belief network to predict osteoarthritic knee pain using data from the osteoarthritis initiative”. AMIA Annual Symposium proceedings. AMIA Symposium, 2008, 788-92.

Webb, G. I., Boughton, J. R., and Wang, Z., “Not so naive Bayes: Aggregating one-dependence estimators”, Machine Learning 58 (1) (2005) 5–24.

Weiss, G. M., and Davison, B. D., “Handbook of Technology Management, H. Bidgoli (Ed.), John Wiley and Sons, 2010.”. 2010.

Wu, P., Manjunanth, B. S., Newsam, S. D., and Shin, H. D., “A texture descriptor for image retrieval and browsing,” Proceedings of The IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'99), Fort Collins, CO, USA, 1999, pp. 3-7.

Xu, K., and Wang, F., “Behavioral graph analysis of internet application”, Proceeding of The IEEE International Conference on Global Telecommunications Conferences (GLOBECOM), pp. 1-5, Dec 2011.

Zaki, M. J., and Meira, JR.W., “Data Mining and Analysis Fundamental Concepts and Algorithms”. (USA,2014).

Zhang, J., and Tan, T., “Brief review of invariant texture analysis methods”. The journal of the pattern recognition society, Pattern Recognition 35, pp. 735–747, 2002.

Zhang, X., Guo, Y., and Du, P., “The Contour Detection and Extraction for Medical Images of Knee Joint”, (2011), Proceeding of The 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, pp. 1-4, 2011.

Zu, Q., and Wenfeng, L., “The research of customer classification based on extended bayes model”. Proceeding of The Third International Conference on Pervasive Computing and Applications (ICPCA), vol 1, pp. 22-25, Oct 2008.

Appendices

Appendix A

Feature Selection Techniques with WEKA software

In this section, the application of WEKA for feature selection techniques for the thesis work are presented. There are five methods are presented: (i) Correlation-based Feature Selection, (ii) Chi-squared, (iii) Information Gain, (iv) Gain Ratio, and (v) ReliefF. The WEKA application of each method is presented below:

1) Correlation-based Feature Selection (CFS)

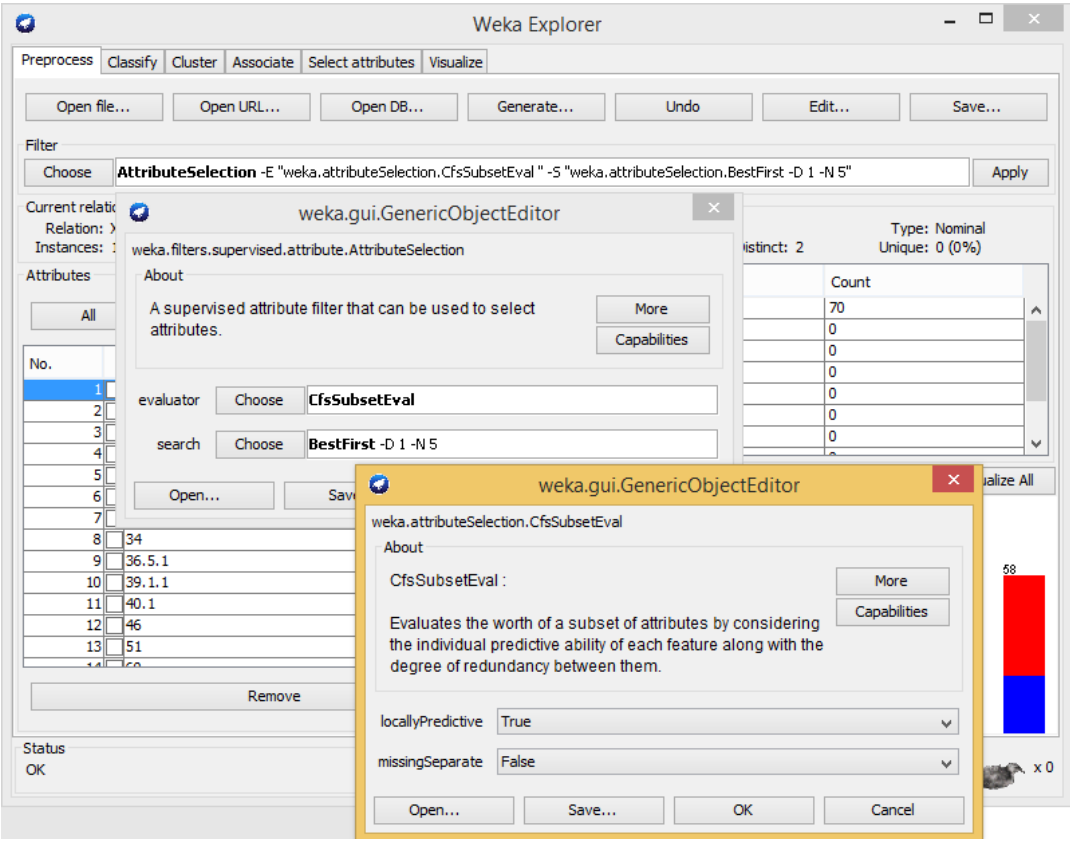


Figure A.1 The CFS Application with WEKA GUI

The Figure A.1 is illustrated the WEKA GUI (Graphic User Interface) application of CFS feature selection technique.

2) Chi-squared

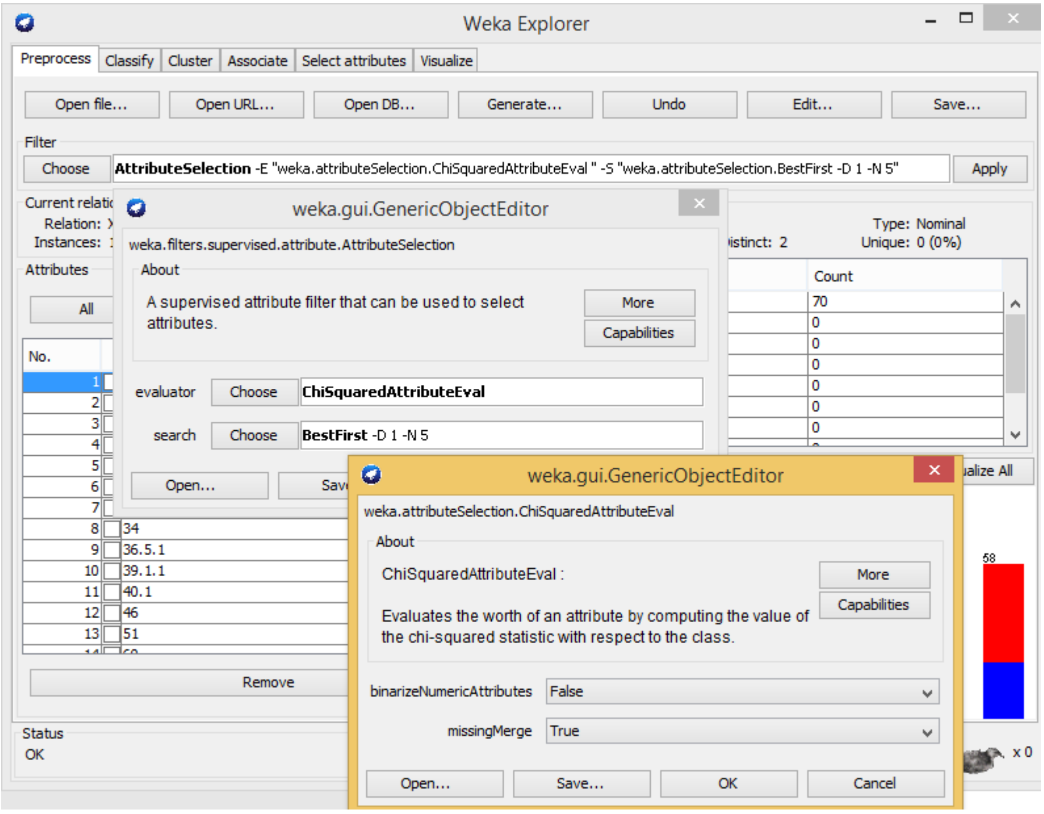


Figure A.2 The Chi-squared Application with WEKA GUI

The Figure A.2 is pictured about the WEKA GUI (Graphic User Interface) application of Chi-squared feature selection technique with the best first ranking.

3) Information Gain

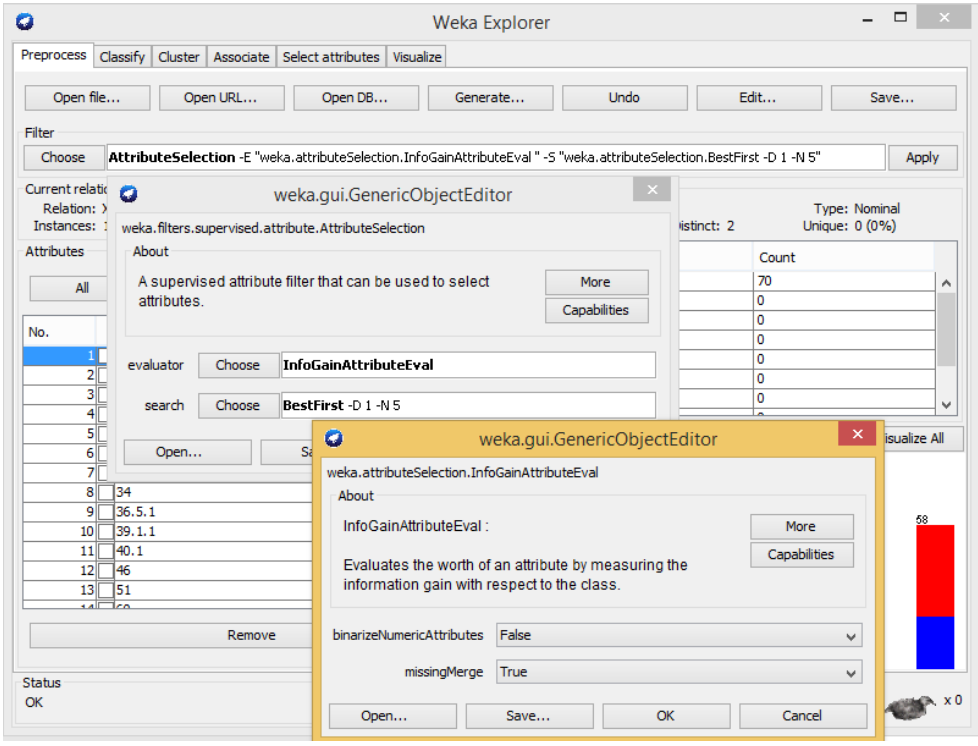


Figure A.3 The Information Gain Application with WEKA GUI

The Figure A.3 is pictured about the WEKA GUI (Graphic User Interface) application of Information Gain feature selection technique with the best first ranking.

4) Gain Ration

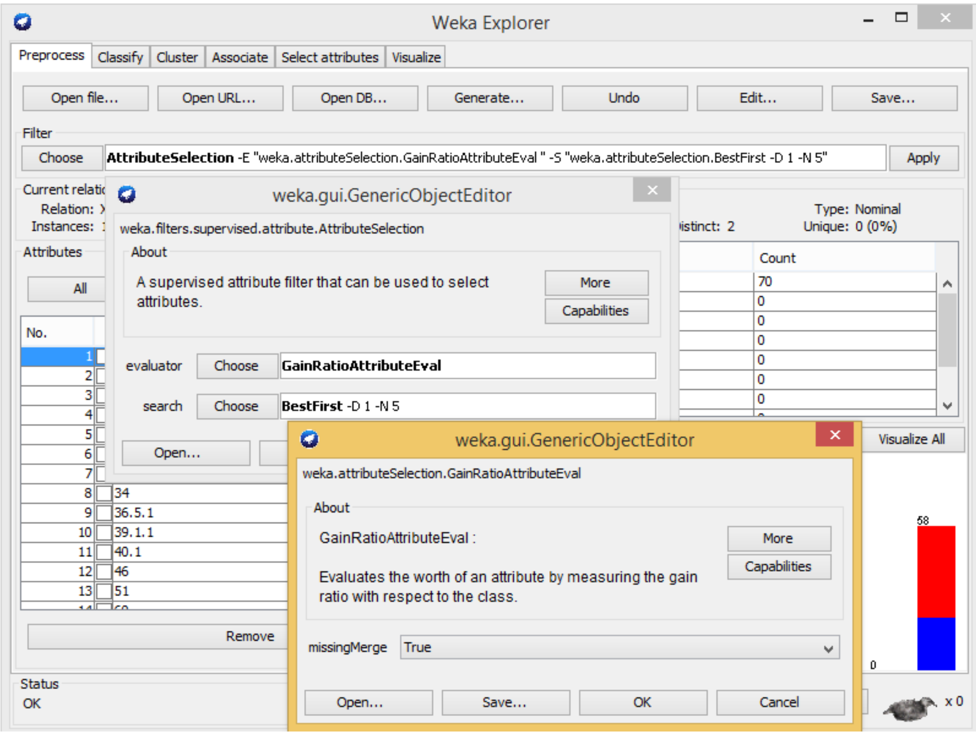


Figure A.4 The Gain Ratio Application with WEKA GUI

The Figure A.4 is illustrated the WEKA GUI application of Gain Ratio feature selection technique with the best first ranking.

5) ReliefF

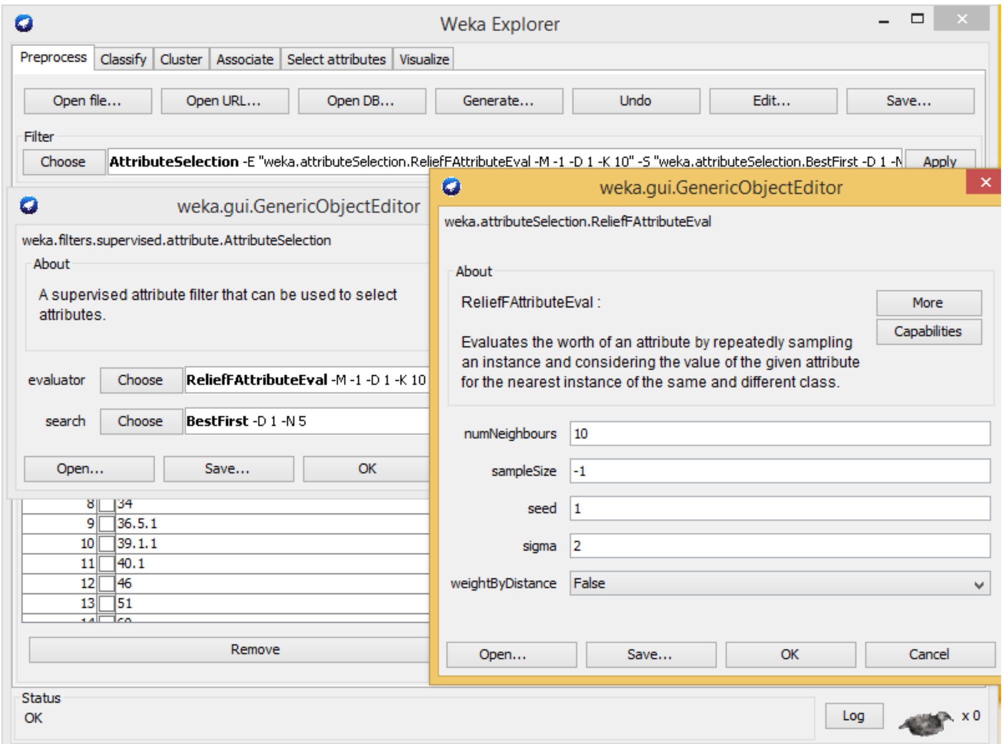


Figure A.5 The ReliefF Application with WEKA GUI

The Figure A.5 presents the WEKA GUI application of ReliefF feature selection technique with the best first ranking.

Appendix B

Machine Learning Techniques with WEKA software

In this section, the application of WEKA for Machine learning mechanisms for the thesis work are illustrated. There are nine methods are discussed: (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, (viii) Sequential Minimal optimization and (ix) Back Propagation Neural Network. The WEKA application of each learning method is presented with each figure below:

1) Decision Tree

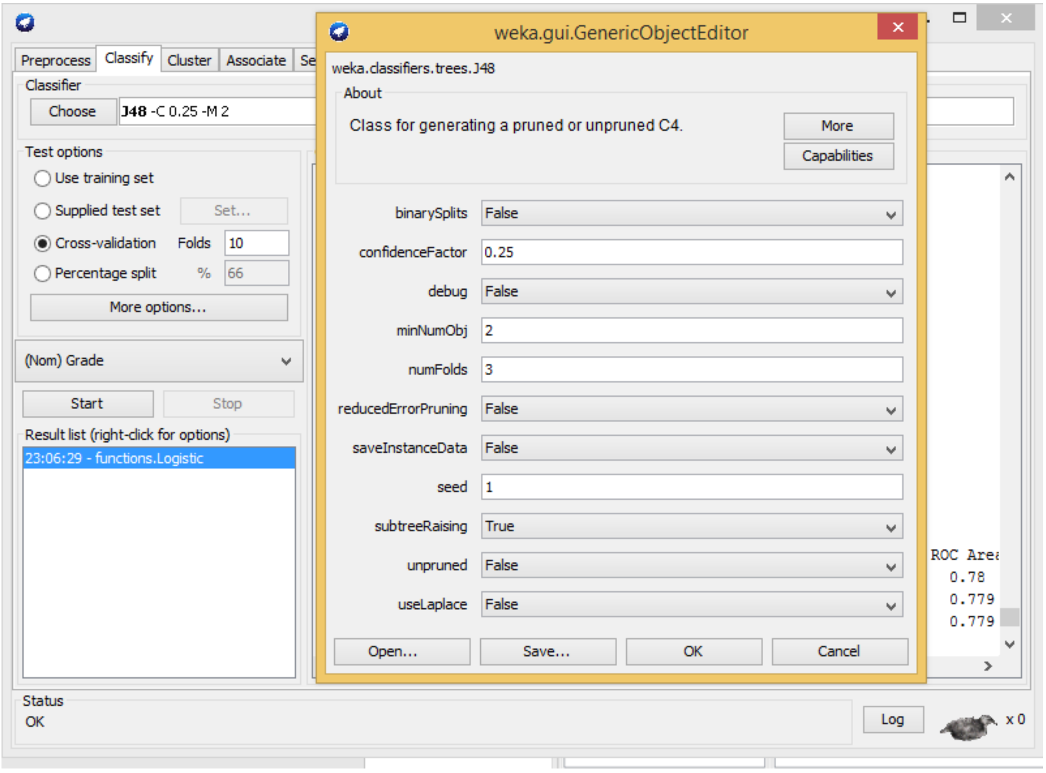


Figure A.6 The Decision Tree Application with WEKA GUI

The Figure A.6 is illustrated the WEKA GUI application of Decision Tree learning classifier.

2) Binary Split Tree

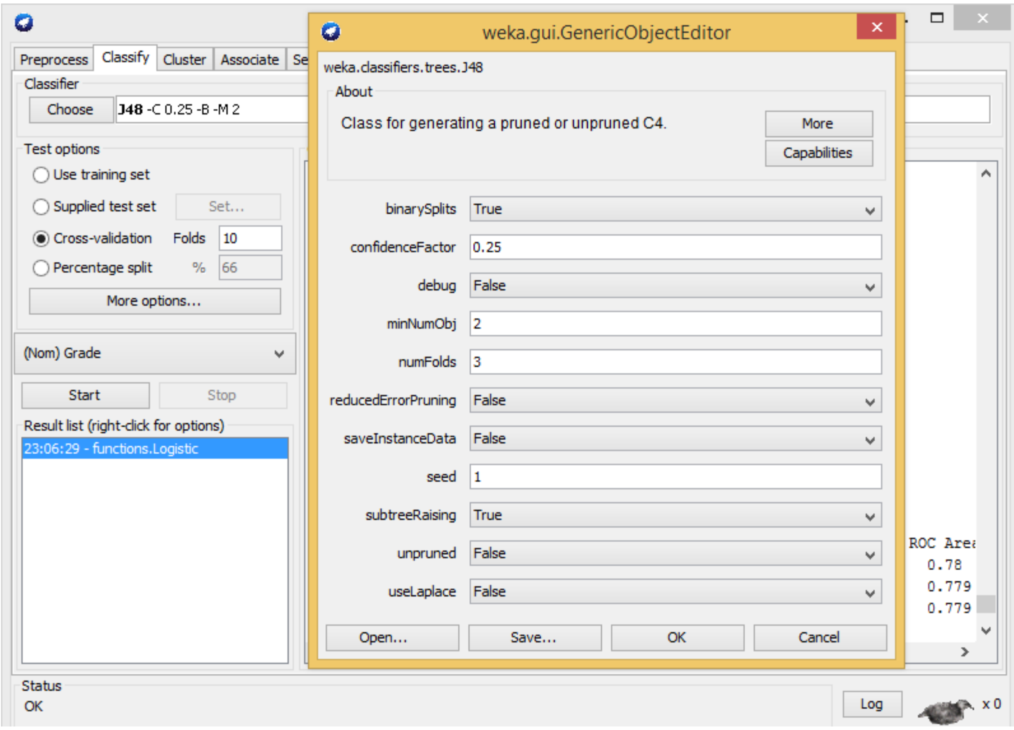


Figure A.7 The Binary Split Tree Application with WEKA GUI

The Figure A.7 presented the WEKA GUI application of Binary Split Tree learning classifier.

3) Average One-Dependence Estimators

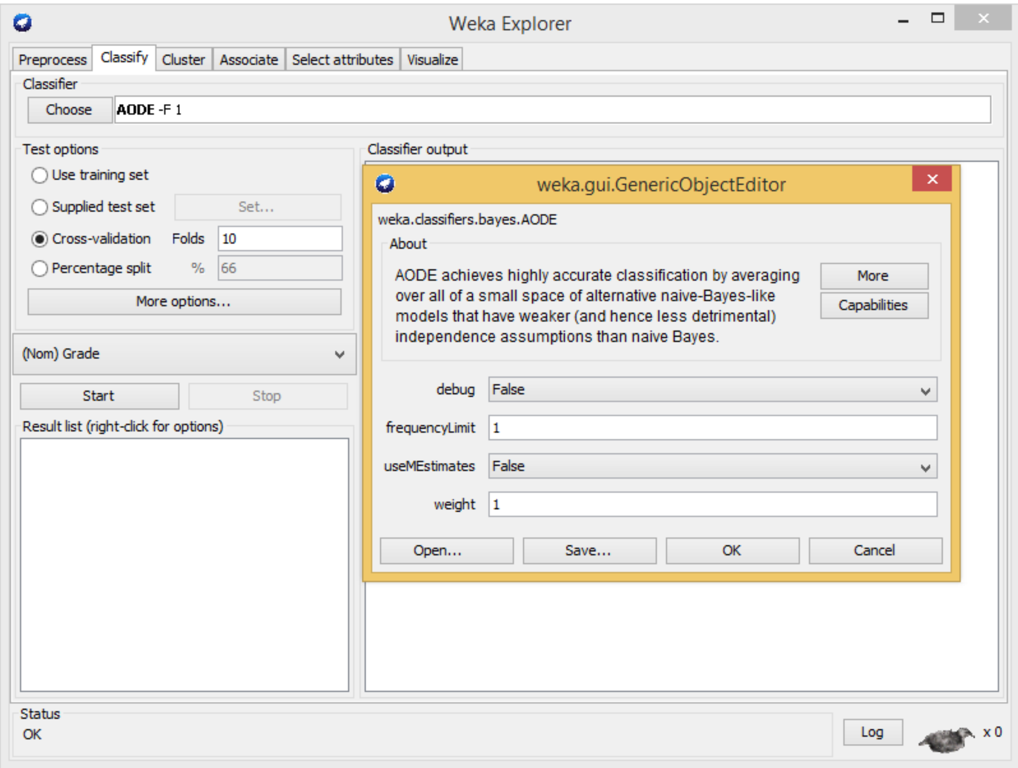


Figure A.8 The Average One-Dependence Application with WEKA GUI

The Figure A.8 shown the WEKA GUI application of Average One-Dependence Estimators learning classifier.

4) Bayesian Network

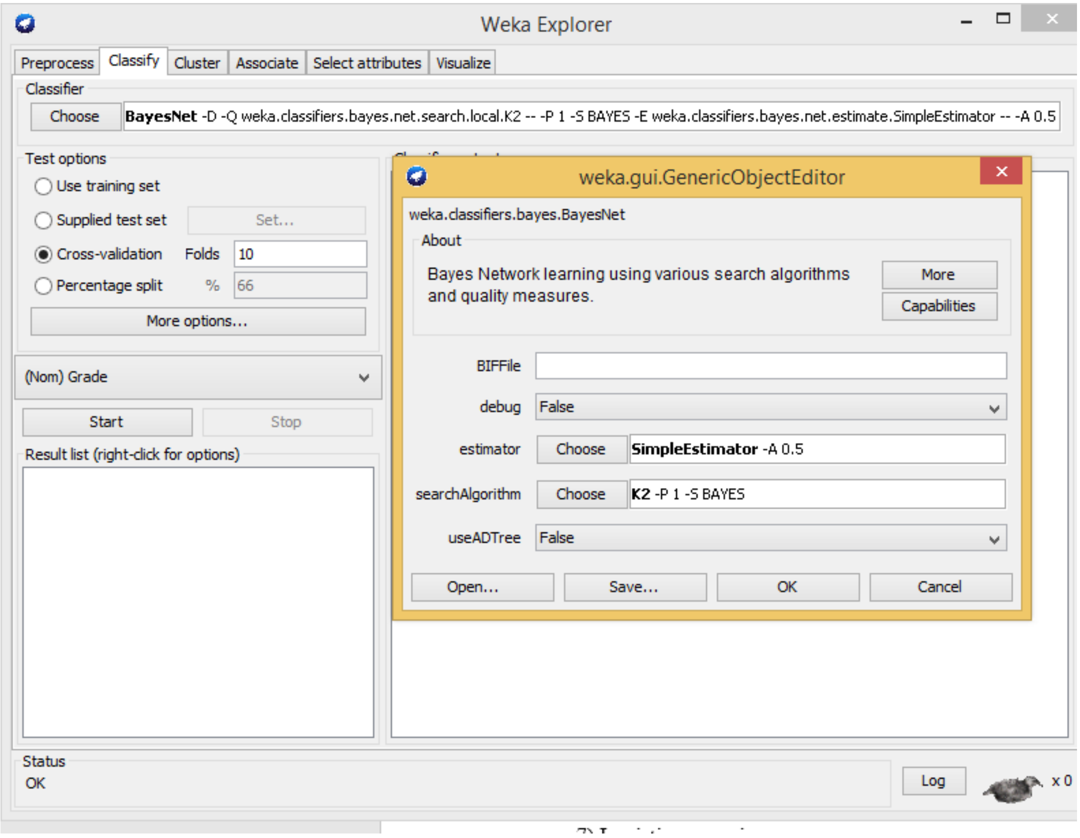


Figure A.9 The Bayesian Network Application with WEKA GUI

The Figure A.9 presented the WEKA GUI application of Bayesian Network learning classifier.

5) Naïve Bayes

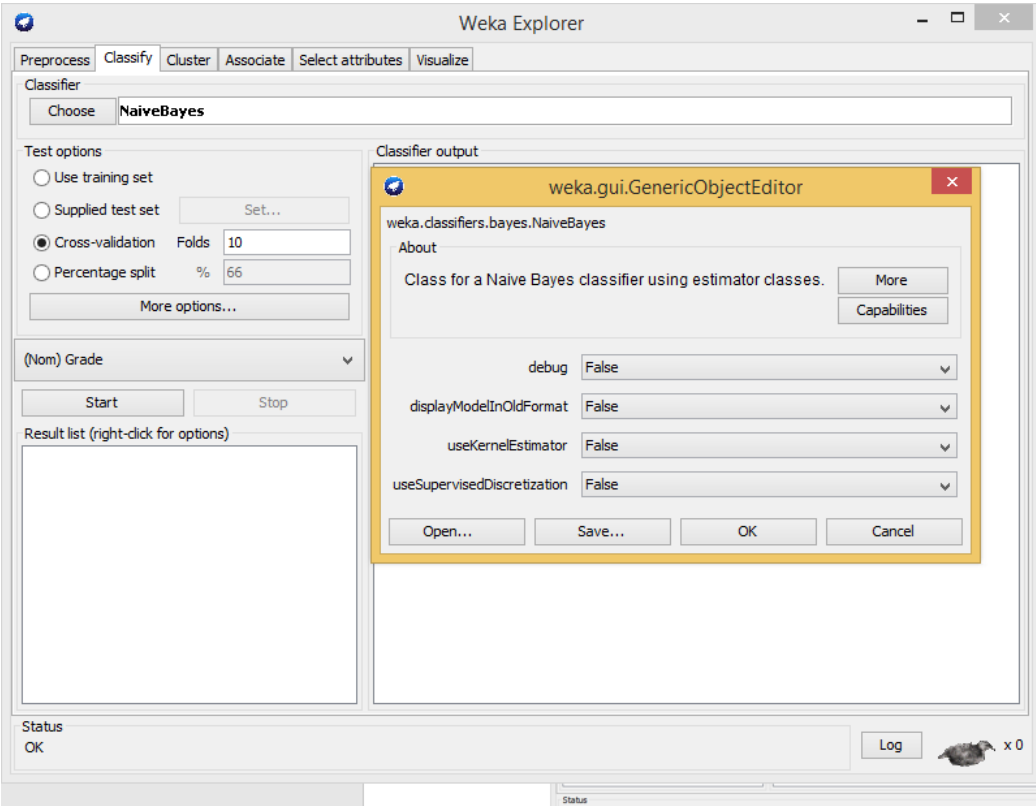


Figure A.10 The Naïve Bayes Application with WEKA GUI

The Figure A.10 illustrated the WEKA GUI application of Naïve Bayes learning classifier.

6) Support Vector Machine

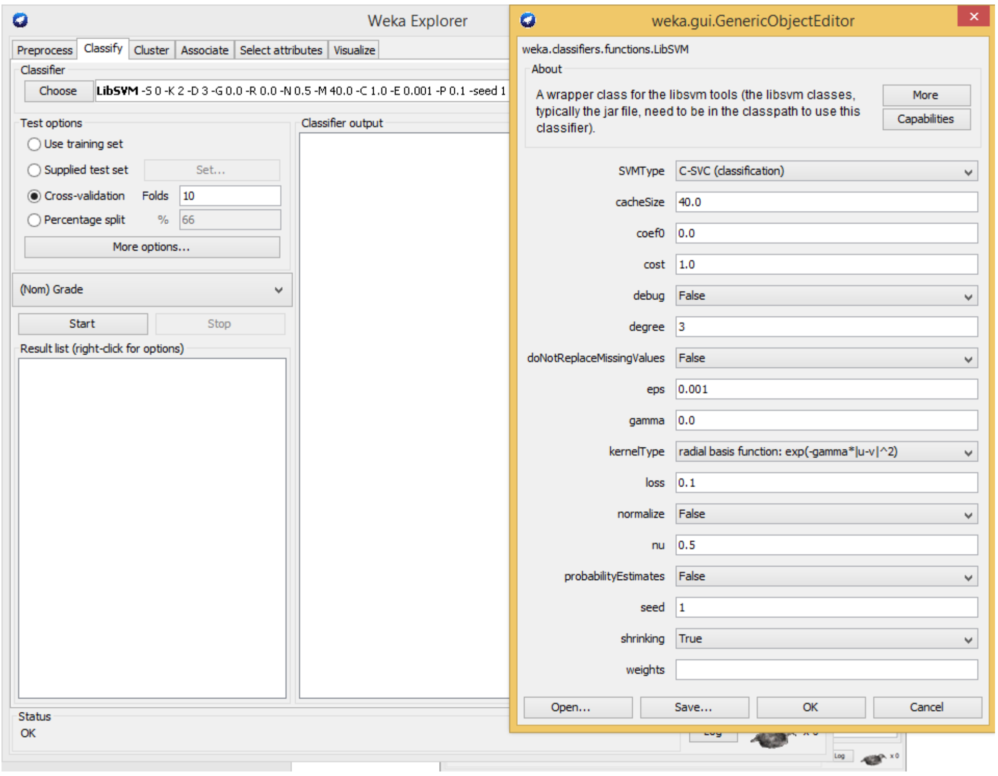


Figure A.11 The Logistic Regression Application with WEKA GUI

The Figure A.11 presented the WEKA GUI application of Support Vector Machine learning classifier.

7) Logistic regression

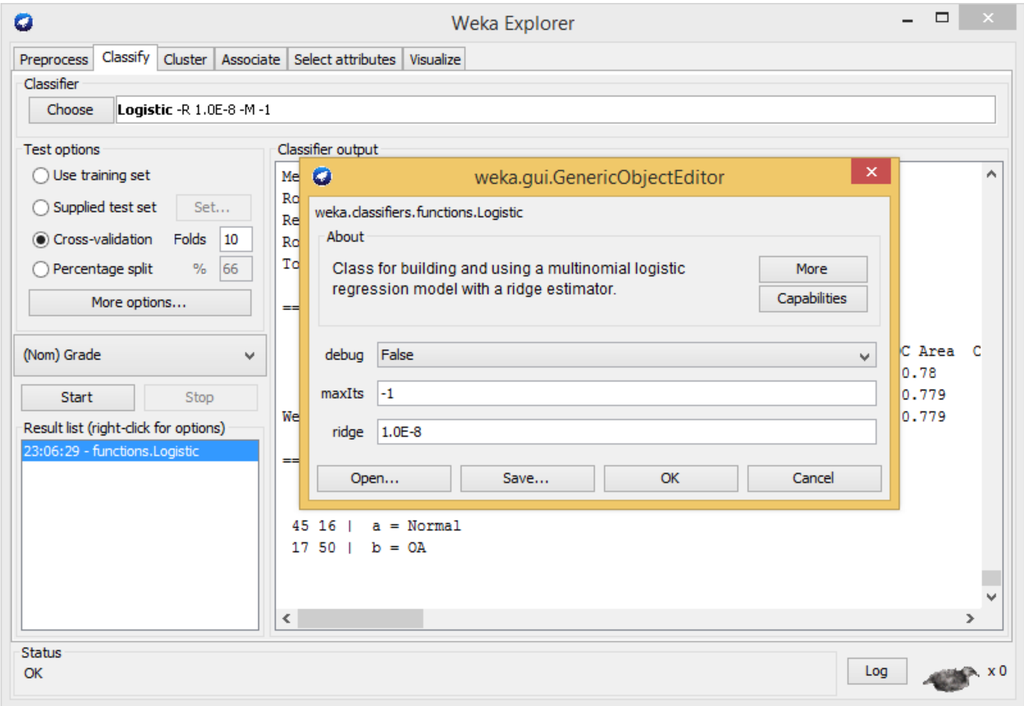


Figure A.12 The Logistic Regression Application with WEKA GUI

The Figure A.12 presented the WEKA GUI application of Logistic Regression learning classifier.

8) Sequential Minimal optimization

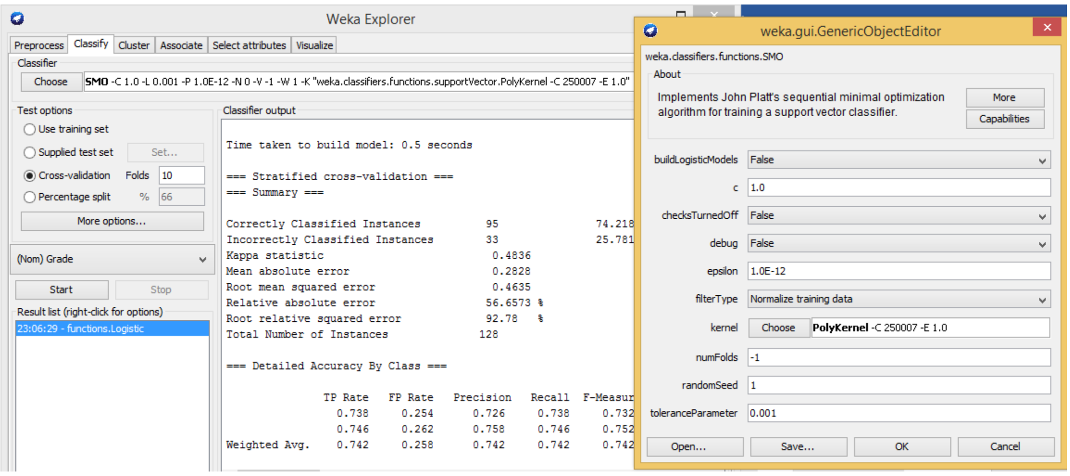


Figure A.13 The Sequential Minimal optimization Application with WEKA GUI

The Figure A.13 illustrated the WEKA GUI application of Sequential Minimal optimization learning classifier.

9) Back Propagation Neural Network

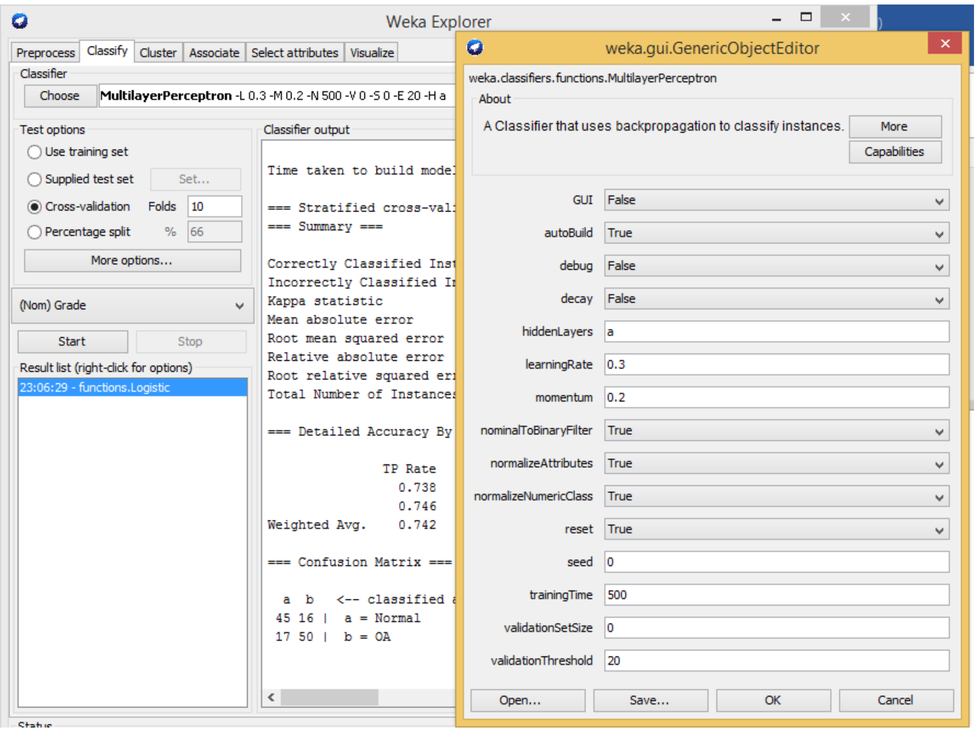


Figure A.14 The Back Propagation Neural Network Application with WEKA GUI

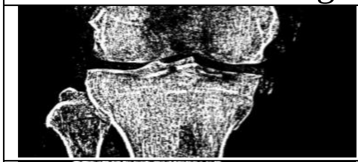
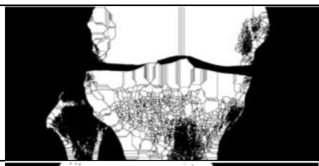
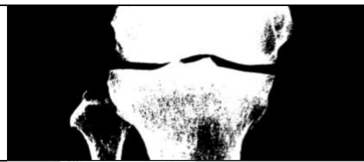
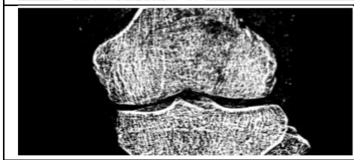
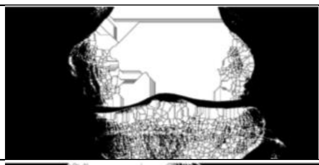
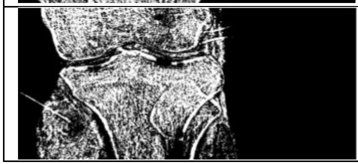
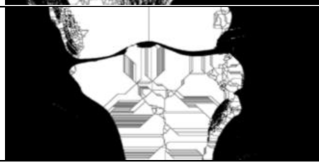
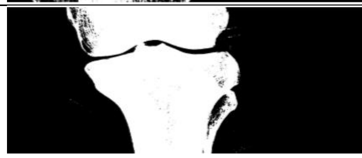
The Figure A.14 is illustrated the WEKA GUI application of Back Propagation Neural Network learning classifier.

Appendix C

Image Enhancement Techniques (Local Threshold, watershed and Otsu)

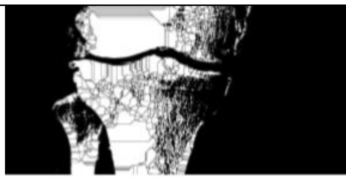
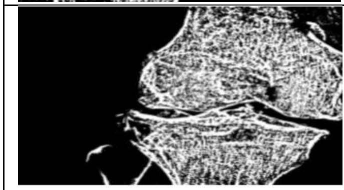
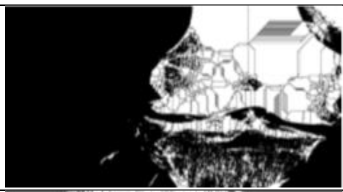
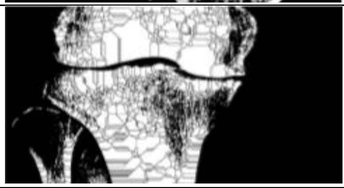
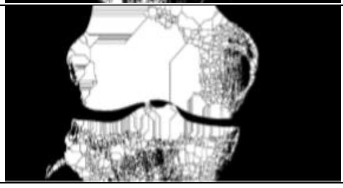
In this section, the comparison of (i) Local Tresholding, (ii) Watershed and (iii) Otsu methods are presented in Table A.1 for normal case and Table A.2 fro OA case imagery.

Table A.1 The Comparison of three enhancement techniques for normal (non-OA) images

Local Thresholding	Watershed	Otsu
		
		
		
		
		

From the Table A.1 presented the three enhancement techniques include: (i) Local Thresholding, (ii) Watershed and (iii) Otsu method. It is suggested that by the medical research field that Otsu give the best performance for make a clear knee.

Table A.2 The Comparison of three enhancement techniques for OA images

Local Thresholding	Watershed	Otsu
		
		
		
		
		

From the Table A.2 presented the three enhancement techniques include: (i) Local Thresholding, (ii) Watershed and (iii) Otsu method for knee OA imagery. It is suggested that by the medical research field that Otsu give the best performance for make a clear knee.

VITAE

Name Mr. Sophal Chan

Student ID 5930223005

Educational Attainment

Degree	Name of Institution	Year of Graduation
Bachelor of Science (Information Technology)	Prince of Songkla University Phuket Campus	2016

List of Publication and Proceeding (If Possible)

Sophal Chan and Kwankamon Dittakan. (2018). "Osteoarthritis Stages Classification to Human Joint Imagery using Texture Analysis: A Comparative Study on Ten Texture Descriptors". *Proceeding of The Second International Conference on Recent Trends in Image Processing & Pattern Recognition*, Solapur University, Maharashtra, India: 21-22 December 2018.