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" [1]. 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 [2] 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 [89,90] studies illustrated that in Asian aging society Knee OA have affected in range from 38.1% to 50.0%.

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 [3, 4, 5]. 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 stages 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 model work powerfully to 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.

(a) Knee without OA   (b) Knee with OA
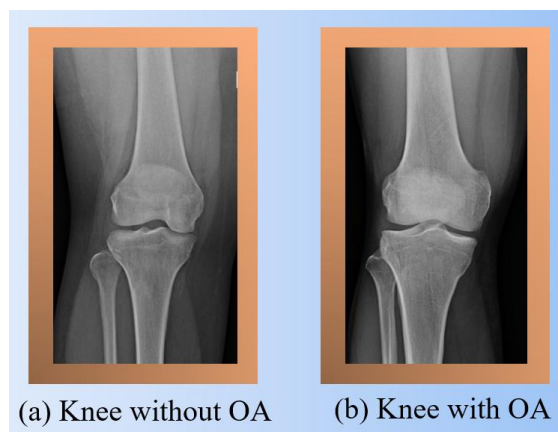
Figure 1.1 Normal and knee OA condition

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 [16,17] 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:

| (a) Normal (Grade 0) | (b) Grade 1 | (c) Grade 2 | (d) Grad 3 | (e) Grad 4 |
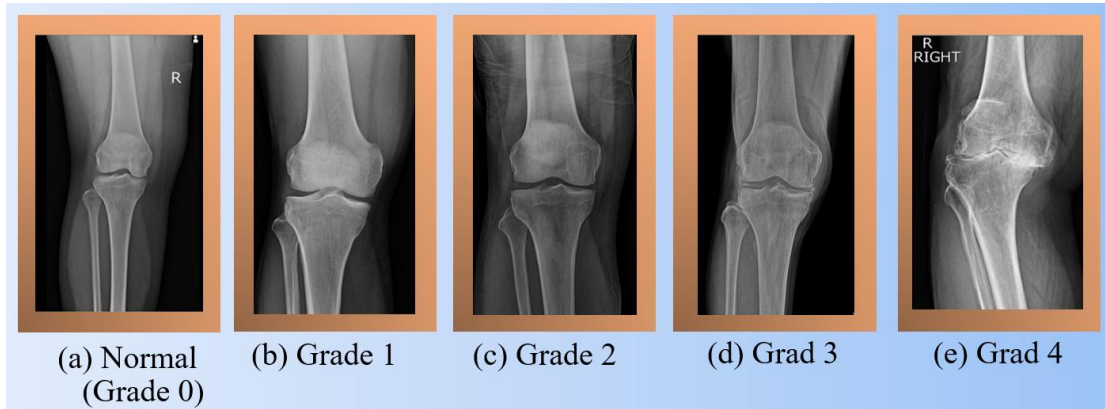
Figure 1.2 The Stage of Knee OA

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 grade 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 graph 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 model 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 model application.

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 model for classification without 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.

7. 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.

I. 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.

II.   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.

III.   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 model 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.