**Chapter 1**

**Introduction and Background**

* 1. **Introduction**

“Knowledge Discovery in Data (KDD) is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data”[86]. 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 within the data. Data mining can be performed on any kind of data repository that includes a relational database [85] where data is stored in a table that has a two dimensional structure with row and column. A data warehouse is a 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 from 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 and medical domains. Extracting useful knowledge from within these images presents a significant challenge. Image mining also encompasses elements from 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. In image classification, a collection of prelabelled images are 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.

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 process building a classification model that can be 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 that this process is the step of building a classifier describing 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 [87, 88, 89]. The classification is used to predict discrete and/or unordered class label. The accuracy of predictive models is defined by using it to pre-label test data.

The reminder 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. **Thesis Objective**

From the forgoing the research concerning at which this thesis work is directed is thus concern with the investigation, realization and evaluation that can be create classification models for the purposed of two studies include knee OA detection study and 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 question:

1. How to obtain the Area 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. **Research Methodology**

To act as a focus for the work a set of medical X-ray were collected from two local hospital: (i) Dibuk hospital (Phuket branch) and Bangkok hospital (Phuket and Suratthani branch), a collection of these X-ray were used in this thesis work. In the collection, there were 2 categories of normal knee and OA image. Figure 1.1 illustrates the normal knee and OA knee X-ray image. This collection was used form knee OA detection study, while the knee OA stages classification study have to need the help from specialist who works at Thungsong Hospital (Nakhon Si Tham Rat) to detects OA based on Kellgren and Lawrence system. The Figure 1.2 presents the stage of knee OA. The data was collected using a sampling process, by collection medial digital X-ray image from the hospital server. For the purpose of the research, the image have to find the region whoch is produce the better performance to classification result that called Region of Interest (ROI).

The first step in the methodology of this work was to segment the appropriate region that is called ROI for forward to the next step. In this ROI segmented, 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 approach for the study: (i) Texture based, (deep learning model, and (iii) graph 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 (ROC), 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 table.

* 1. **Contribution**

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

1. A Comparative Study of Texture Analysis Techniques for Osteoarthritis Classiﬁcation Using Knee X-ray Imagery (Section 4.2 of Chapter 4).
2. Osteoarthritis Stages Classiﬁcation to Human Joint Imagery using Texture Analysis: A Comparative Study on Ten Texture Descriptors (Section 4.3 of Chapter 4).
3. Convolutional Neural Network Application to Medical X-ray imagery for detecting Knee Osteoarthritis by applying Alex Network pre-trained model of transfer learning (Section 5.2 of Chapter 5).
4. The Application of Transfer Learning in Convolutional Neural Network for Osteoarthritis Stages Classification to human joint Imagery using Alex Network Pre-trained Model (Section 5.3 of Chapter 5).
5. Knee Osteoarthritis Detection to Human Joint Imagery using Quadtree Decomposition of graph based approach to Medical X-ray imagery (Section 6.2 of Chapter 6).
6. Graph-based Approach with Quadtree decomposition Application of Knee Osteoarthritis Stages Classiﬁcation to Human Joint Medical X-ray imagery (Section 6.3 of Chapter 6).
   1. **Publications**
   2. **Thesis organisation**

The arrangement of the reminder of this thesis is as follow. Chapter 2 provides an extensive literature review image analysis in term of knee OA detection and the previous work concerning the technologies that feature in tis thesis. Chapter 3 illustrated the way of selected ROI for each learning approach in order to apply for knee OA detectionand the ways to reduce unuseful images for the study work. For Chapter is illustrated the stage of knee OA which applied with texture based approach. The application of convolutional neural network deep learning model for stage of knee OA classification illustrate in Chapter 5. For graph approach of quadtree implementation to detect the knee OA stages presents in Chapter 6. Finally, the thesis concluded with the summary, presentation and discussion of the main findings in term of the research question and sub-question defined above and some direction for future work.

* 1. **Summaries**

This chapter have pictured 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.