**Chapter 2**

**Background and Literature Review**

**2.1 Introduction**

In this chapter the background and the some previous works are described. The content of this chapter is organized as follows: Section 2.2 Medical Image, a brief history of medical image technology and the implementation of medical image in human debases diagnose, typically Knee-OA is presented. Section 2.2 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 2.2.1, Sub-section 2.2.2 and Sub-section 2.2.3, respectively. The summary information of image classification is suggested in Section 2.3. 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 2.3.1 to 2.3.3, respectively. The process after get the classification result the main problem is how evaluate the result. Thus, the evaluation measurement methods is discussed including: (i) Confusion Matrix, (ii) Five evaluation measures (Accuracy, Recall, Sensitivity, Specificity and F-Measure), (iii) Area Under Curve (AUC) and (iv) Statistical model used to analyse the differences across multiple test attempts are presented in Sub-section 2.4.1 to 2.4.4, respectively. Finally, the summary of this chapter will be described in section 2.5.

**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) Ultra-sound, (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 are 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**

In 1895, a German physicist Wilhem Rontgen discovered X-ray image. The first medical image of X-ray image was made from his wife’s hand. Wilhelm Rontgen took this radiograph of his wife's left hand on December 22, 1895, shortly after his discovery of X-rays. The image of Wilhelm’s wife’s hand is shown in Figure 1.

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 1 The image of Wilhelm wife left hand [32].

There are five types of principal radiographic densities in X-ray image. The Figure 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 metal (the white/ the lightest area) [33].

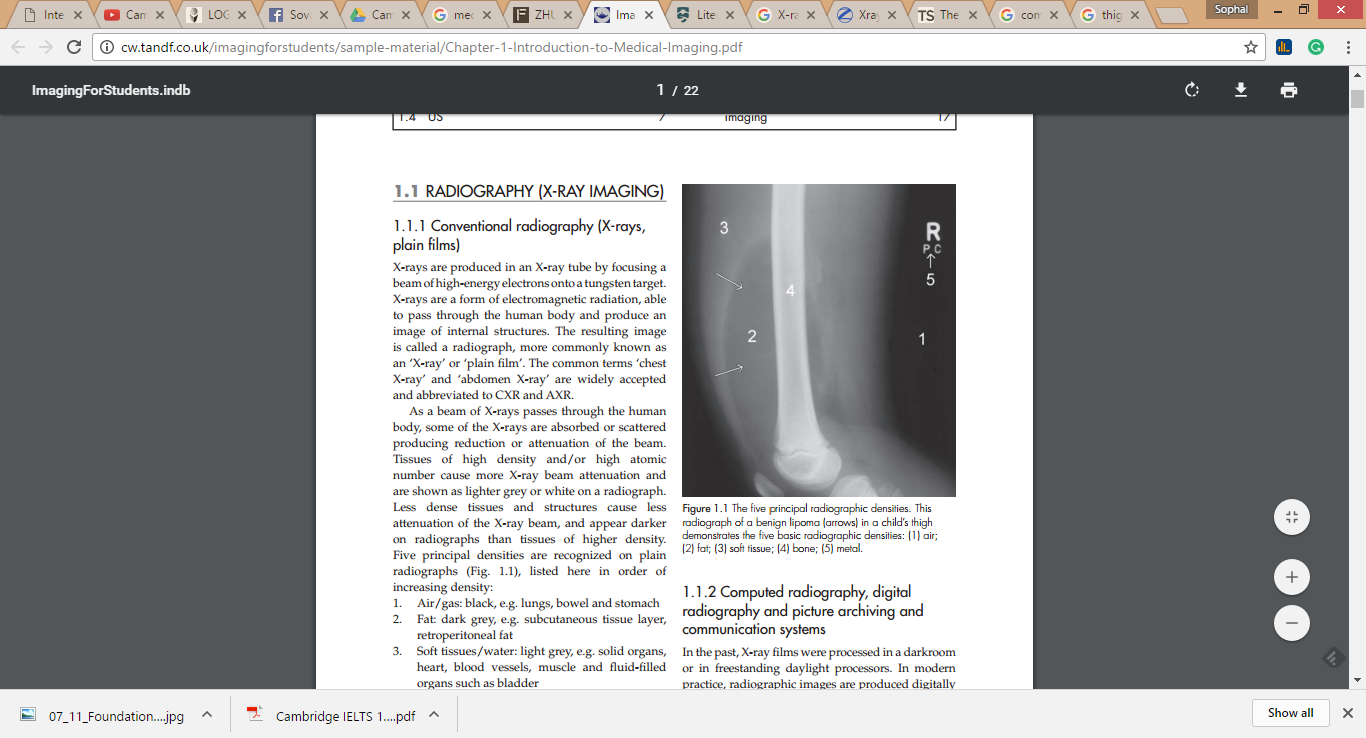


Figure 2 The difference radiographic density in X-ray image [33].

Medical imaging is the technique and process of creating visual representations of the internal structures hidden by the skin and bones inside of a body such as bone, brain, lung and liver which used for medical analysis, as well as to diagnose and treat disease.

X-ray images were used for diagnose Knee-OA which presented in many works [1,2,3,4,5,6,7,8]. There are two main objectives for Knee-OA research: (i) first the X-ray image was used to detect OA and non OA [3,1], while the second one is more advance than the first one because the X-ray imagery was used to detect the OA stages [2,4,5,6,7].

On the purpose of the detection of the OA stage, the Kellgren-Lawrence (KL) system [9,10] is considered. The KL standard is a five levels indicating using 0 to 4. ‘0’ represents no OA symptom while ‘4’ represents the most serve radiographic disease. The summary of the OA stage is shown in Table.1. There are five stage of OA has shown in the table which rank from zero to fourth stage. The stage 0 refer 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 which consider 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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **OA Stage** | **Stage 0**  **(normal)** | **Stage 1**  **(doubtful)** | **Stage 2**  **(minimal)** | **Stage 3**  **(moderate)** | **Stage 4**  **(severe)** |
| **Feature** | 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 |

Table 1 The summary of the OA Stages.

Figure 3 shows the X-ray image of each OA levels: Figure 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.

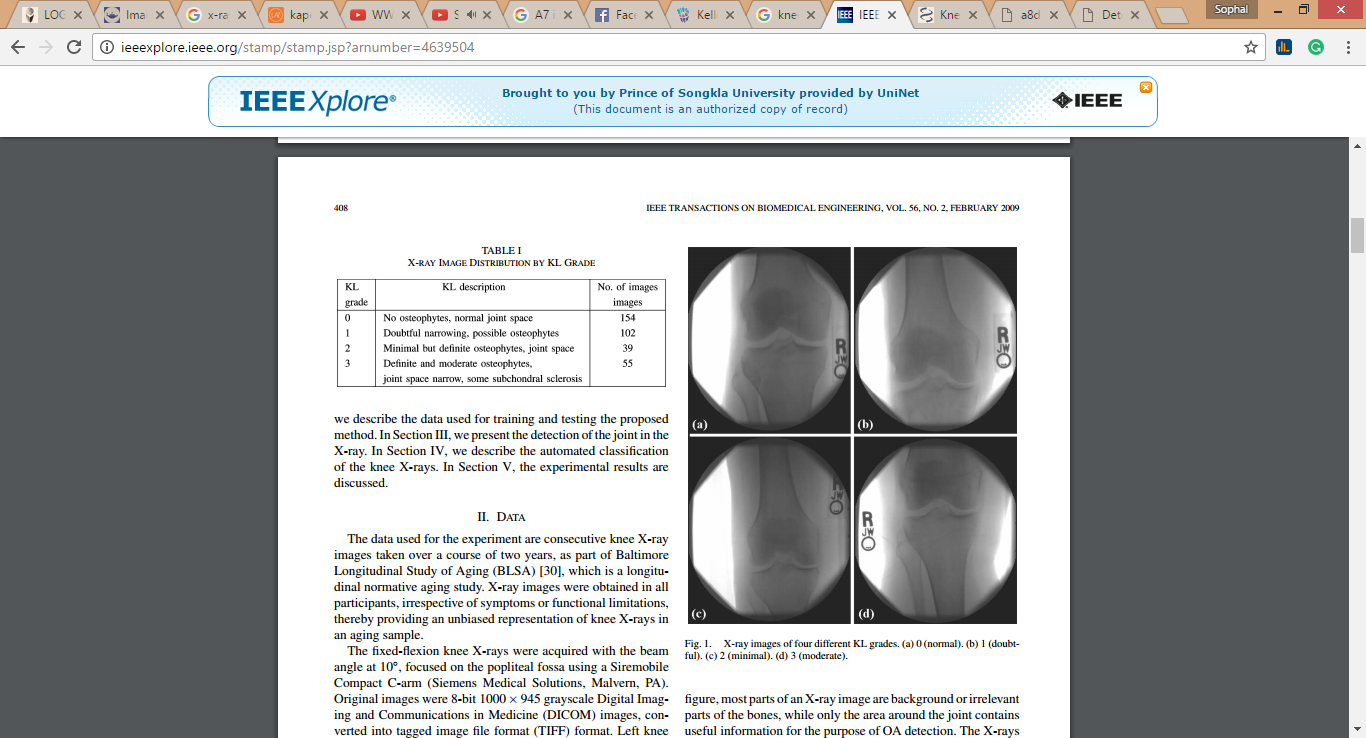


Figure 3 X-ray image of four different KL grade [34].

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 medical 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 unexpansive compare to the others.

**2.2.2 Computed Tomography**

Computed Tomography (CT) is one of the well-known medical imagery 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 about internal structure and very clear image. With respect to OA detection, the MRI images were also applied presented in [11,12,13,14]. 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 4.

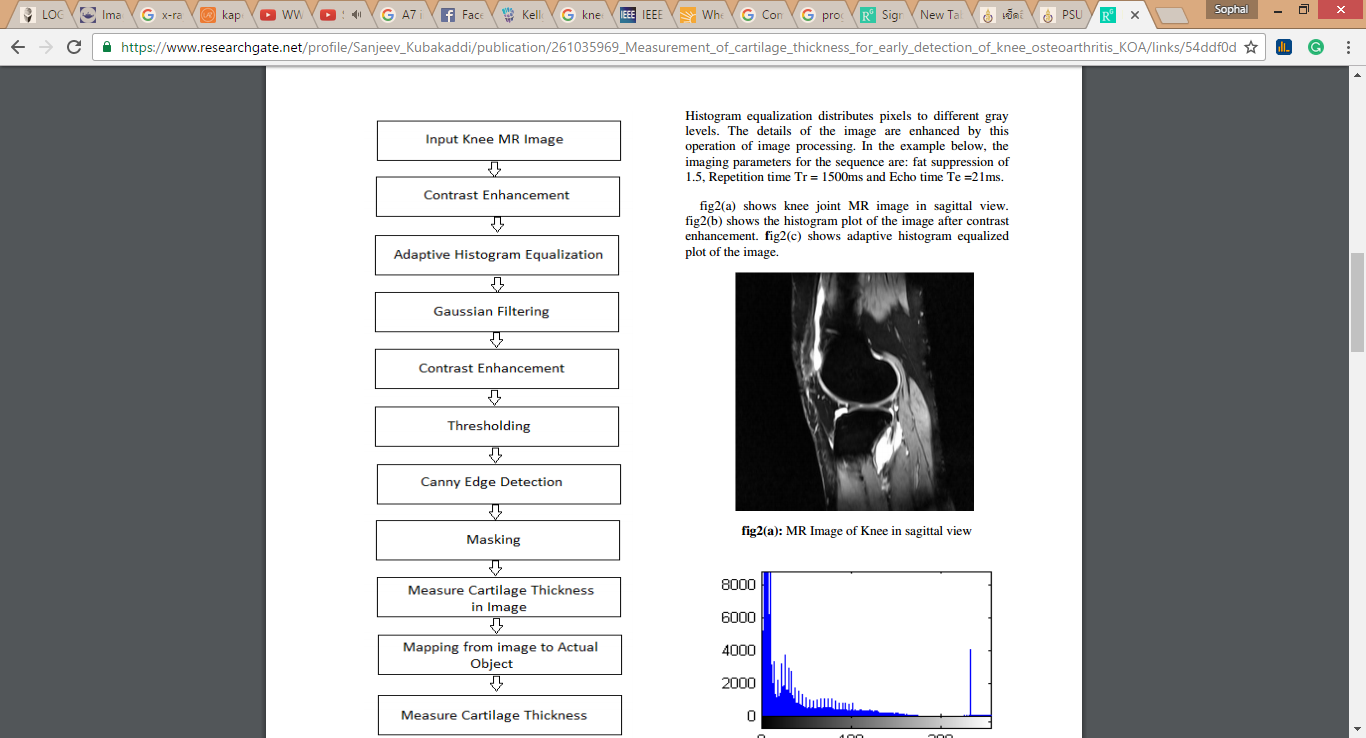


Figure 4 Knee MRI image.

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 medical X-ray image apply with image processing to diagnose OA stage, the image processing technique is presented in the next section.

**2.2 Image Processing**

The fundamental concept image processing is discussed in this section. Image processing is a challenge area in computer science, which has the main purpose to analysis and manipulation of a digitised 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 (ROIs). The smallest element of image is pixel. Image processing generally separate into two main categories: (i) digital image processing and medical image processing [36].

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 5 shows each pixel of greyscale image with the value from 0(black) to 255 (white) because 8 bit equal to 256 grayscales.

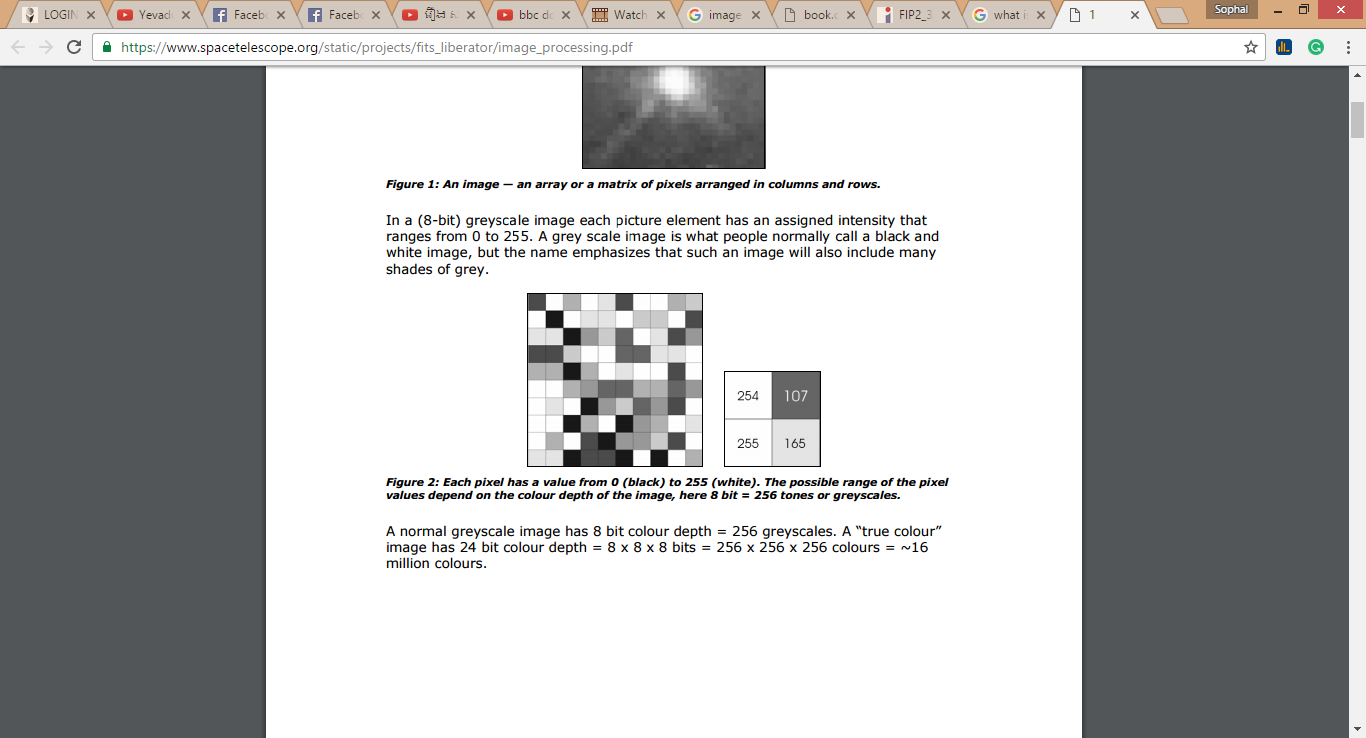


Figure 5 Each pixel has a value from 0 (black) to 255 (white) [35].

There are three image processing processes including: (i) image segmentation, (ii) image enhancement and (iii) image representation/Feature extraction. The detail of each process is discussed in the Sub-section 2.2.1 to 2.2.3 respectively.

**2.2.1 Image Segmentation**

Image segmentation is one of the most important processes of the image processing. Image segmentation is the division of an image into sub-image that is called region which correspond to different object. The main purpose of image segmentation is to divide the given image in to salient image regions, meaningful region, and homogeneous with specific cluster pixel. Moreover, object recognition, occlusion boundary estimation within motion or stereo systems, image compression, image editing, or image database look-up should be use in image segmentation field. 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 applied with based on the different kind of work. Some of existing image segmentation methods for example: (i) Segmentation based on threshold segmentation or clustering, (ii) segmentation based on region and (iii) segmentation based on edge. The brief detail of each technique is discussed in this sub-section.

T**hreshold segmentation**

Threshold segmentation is one the existing methods for image segmentation. The separation of light and dark region is the primary purpose of threshold segmentation. In general, the Thresholding is applied to create the binary image which make from grey-level ones by turning all pixels that below threshold become zero (0) and all pixel above threshold 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 belows:

I(x,y)=………………………(1)

Threshold 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.

There are many applications using threshold mechanisms were proposed for medical image segmentation. In [15] 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 was high complexity of the computation. Figure 5(a) shows the original of human brain scan and Figure 5(b) presents the corresponding segmented image using Otsu method.

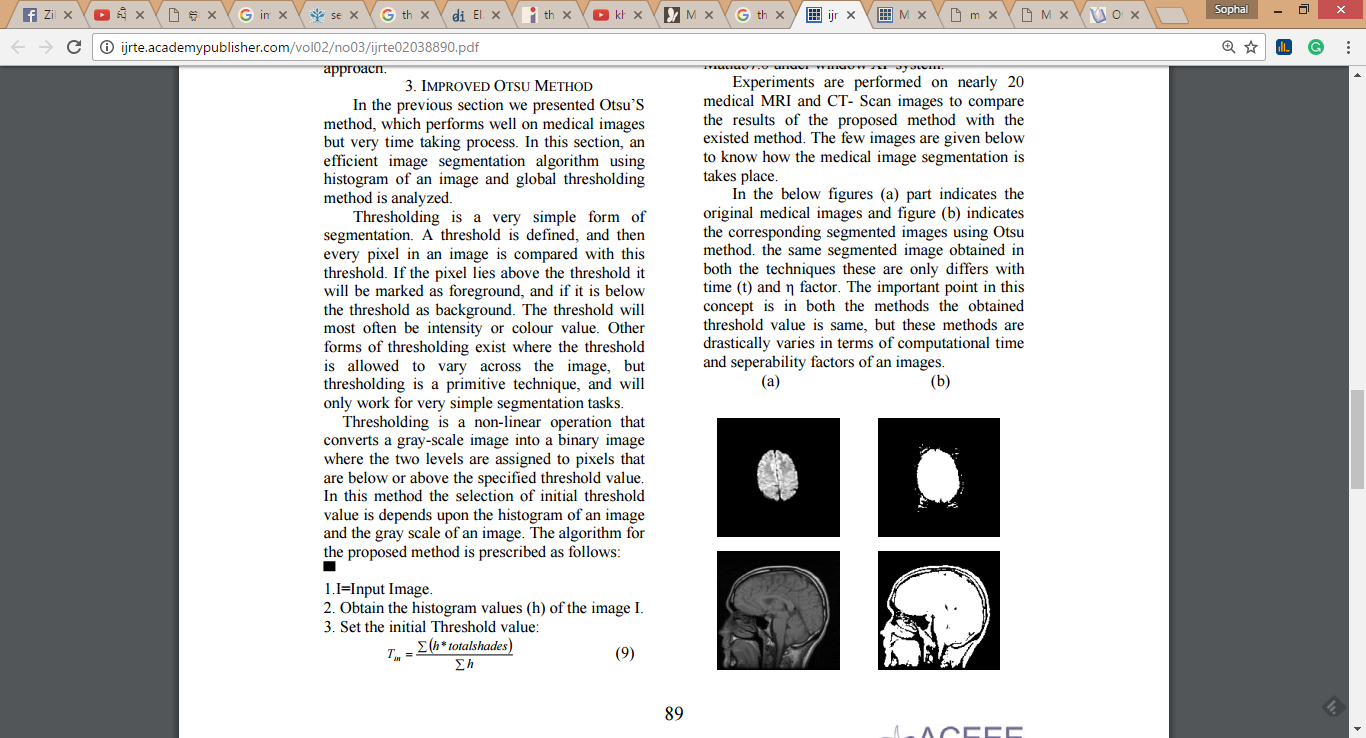
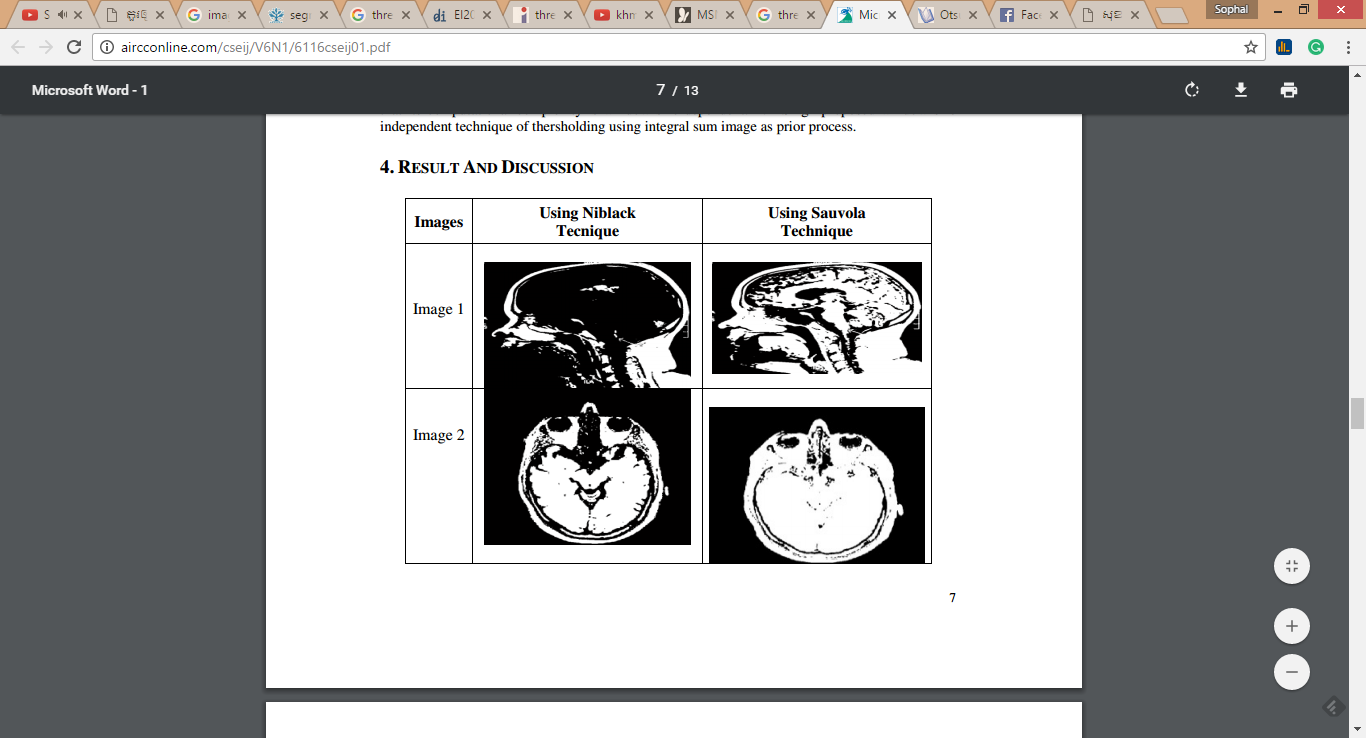
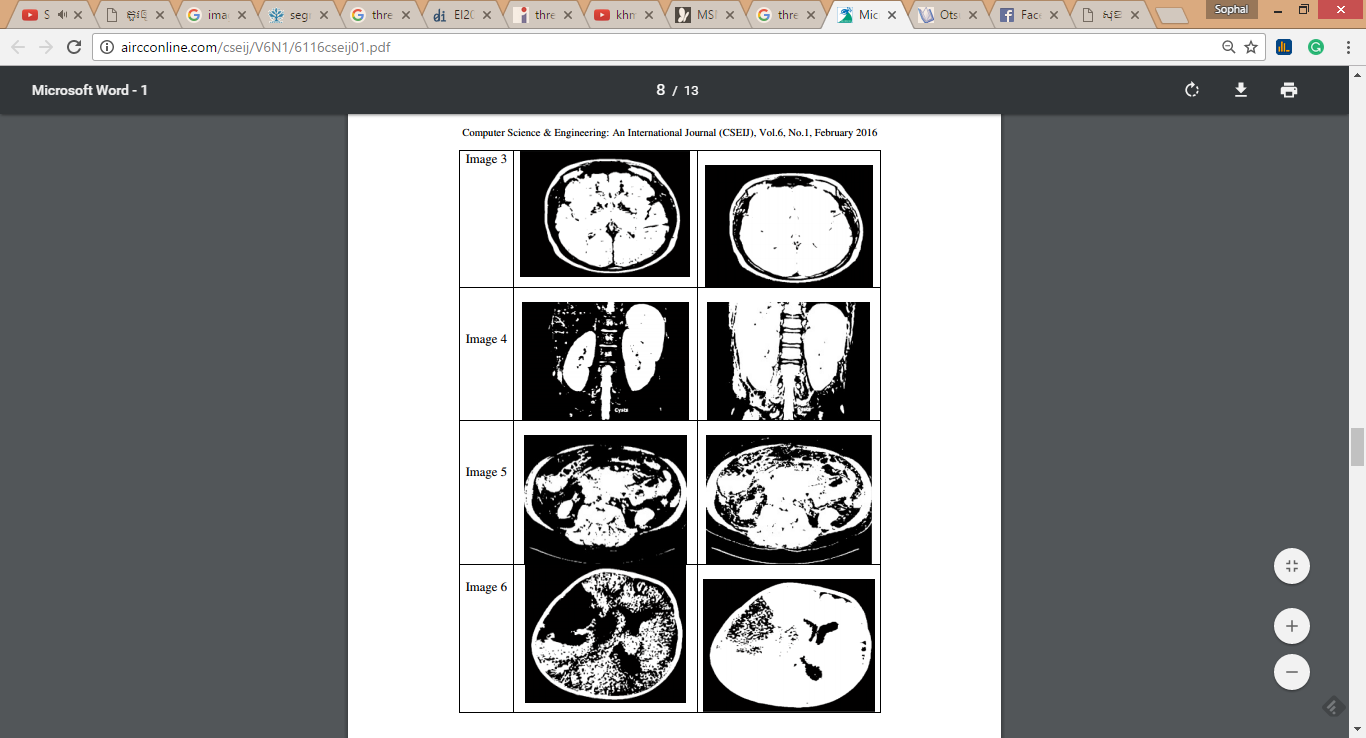
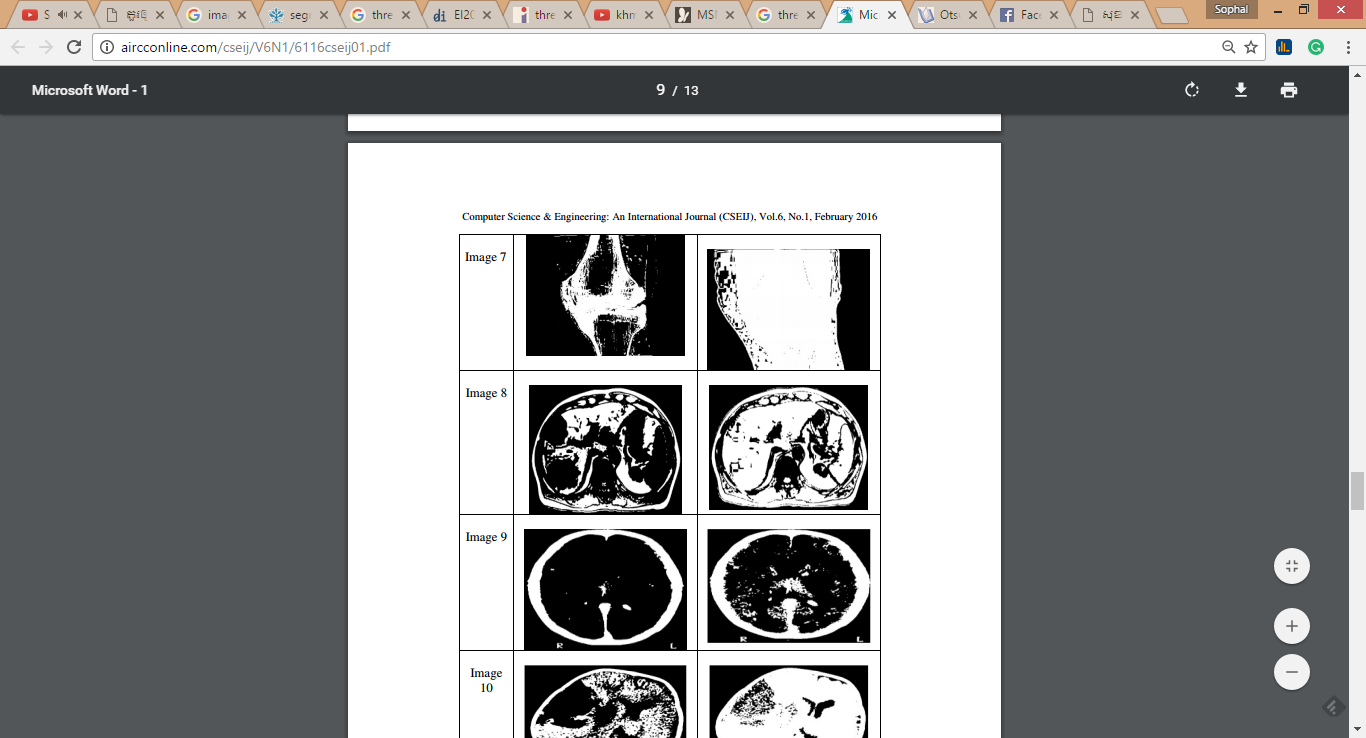


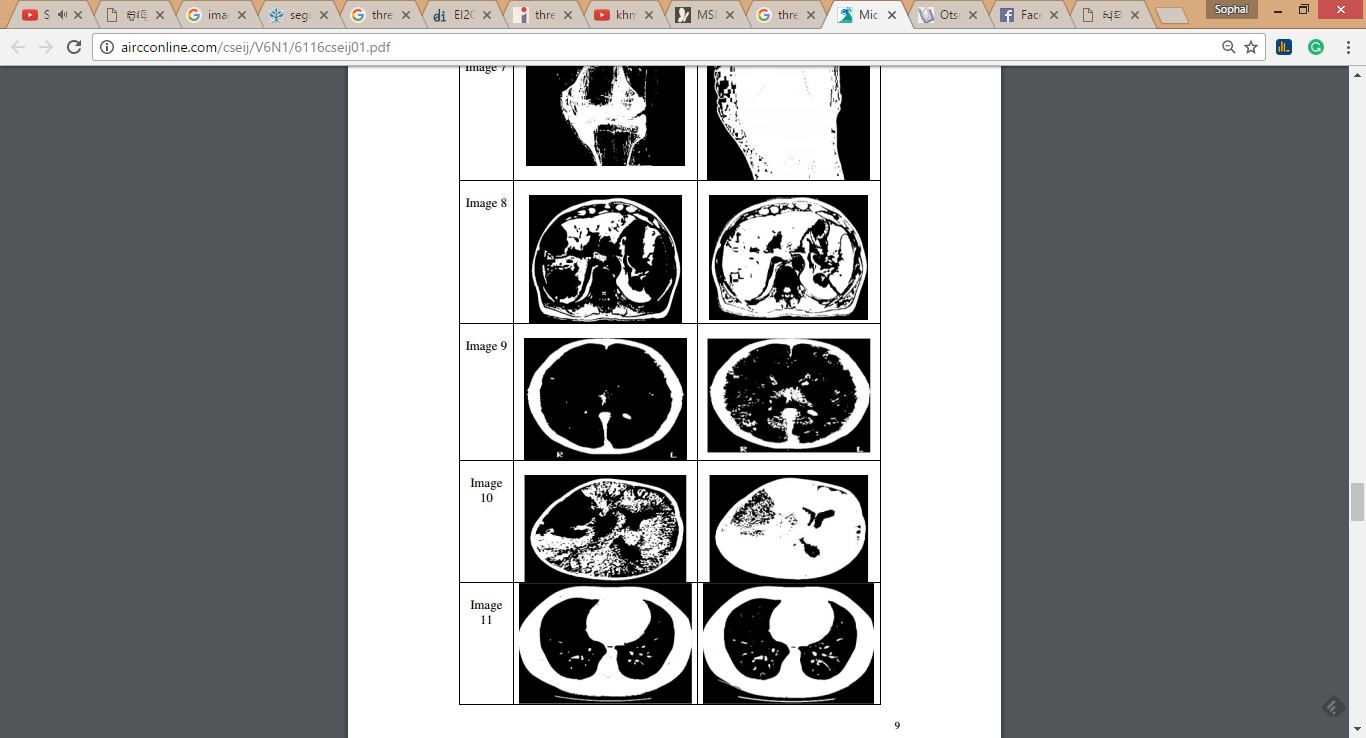
Figure 6 The human brain scan using Otsu method (15)

Local adaptive thresholding also is applied in [16]. The work reported the implementation of local adaptive threshold technique to removes background using local mean and standard deviation. The Niblack and Sauvola local thresholing algorithm were used for comparison. The results shown that the Niblack local threshold algorithm reduced the background noise better than Sauvola local threshold algorithm. Figure 7 shows the difference of twelve sub-images of Niblack technique and Sauvola technique.









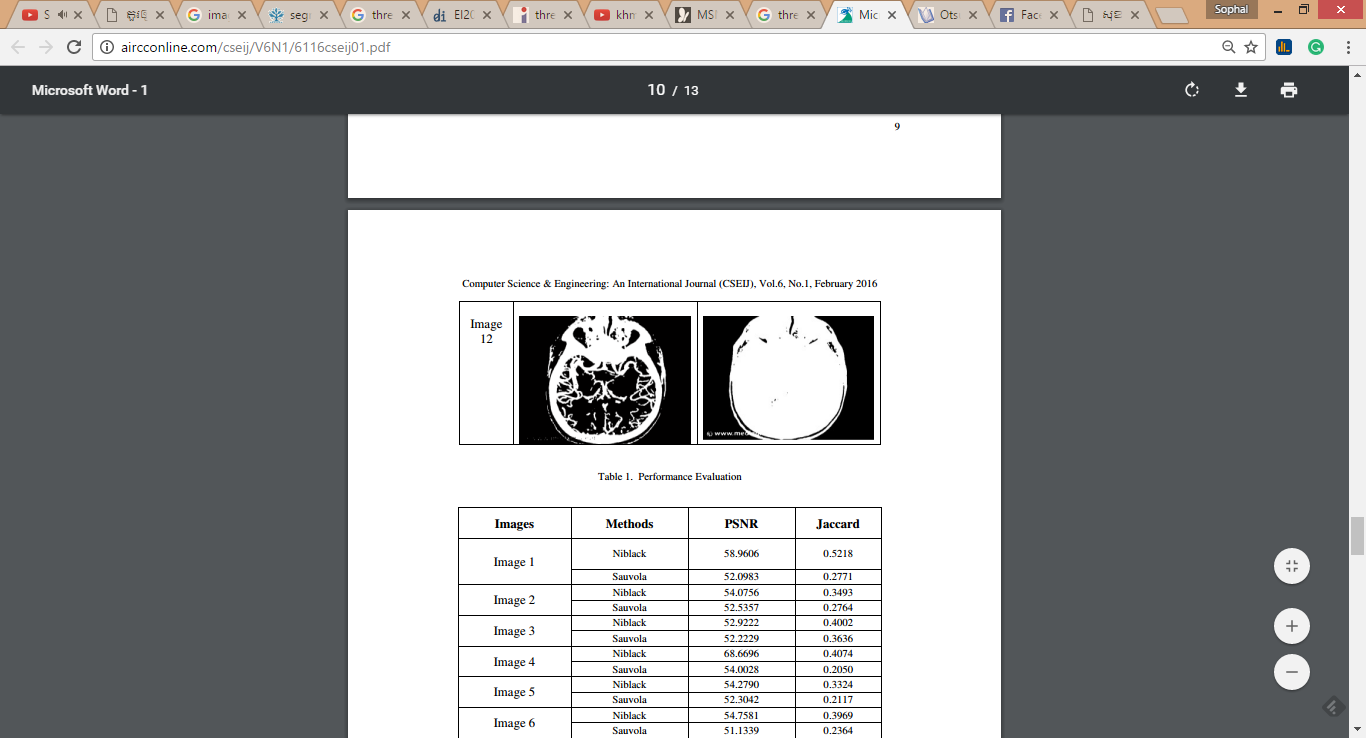


Figure 7 The comparison of 12 medical image by using Niblack and Sauvola Technique [16].

The local adaptive thresholding and Otsu thresholding method are the two techniques of many well-known techniques in medical image segmentation.

**Region-Base Segmentation**

Region-based segmentation is one of image segmentation techniques which has the main goal to find coherent regions in the image. The coherent region contains pixels which 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 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 [17]. The work presented in [17] reports 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 8.

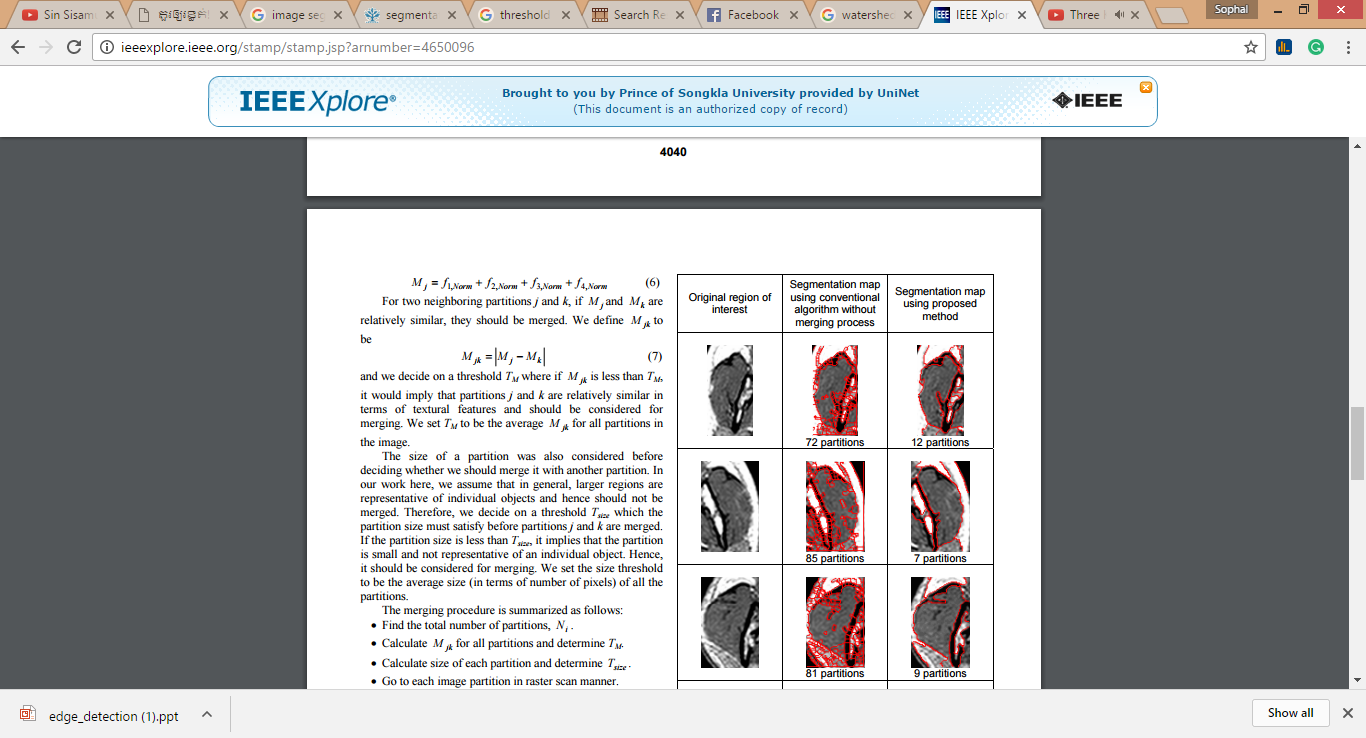


Figure 8 The comparison of image using watershed segmentation

with/without texture-based [17]

**Edge-Base Segmentation**

In edge-base 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-base 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 9 shows the Prewitt and Sobel operator.

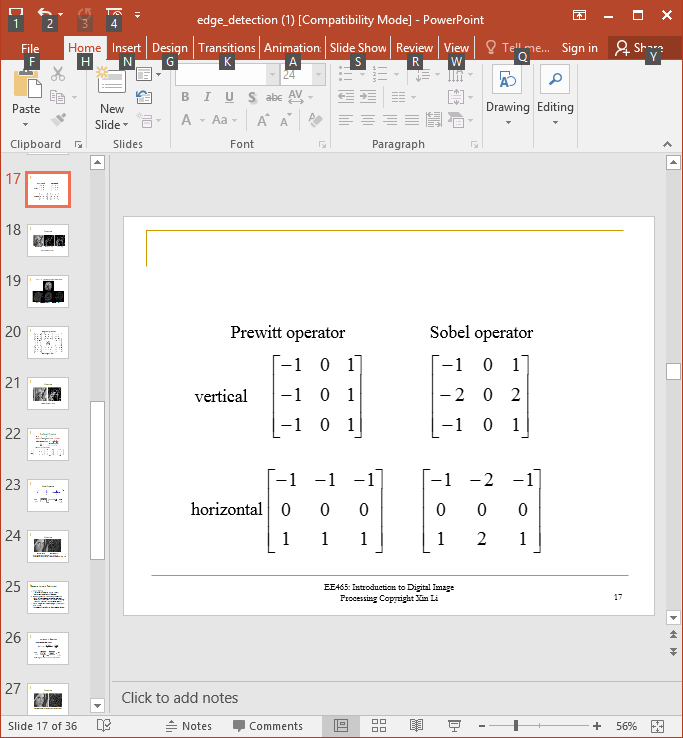


Figure 9 Prewitt and Sobel operator for edge detection

With the respect to the edge detection in the context of medical image segmentation, it can be found in [18]. 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 10.

|  |  |
| --- | --- |
|  |  |
| (a) Original brain scan image | (b) The processed image using edge detection |
|  |  |

Figure 10 The examle of brain scan image using edge detection[18]

From Figure 10(a) shows the original brain scan while Figure 10(b) illustrates the brain scan image after the edge detection was applied.

**2.2.2 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 varies type of image such as personal images, medical images, satellite images and aerial images. The photographer/user/viewer 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. Typically, the image enhancement is applied is to improve the quality and perception of information contained in the image for human viewers. During image enhancement process an input is an 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 factors for example hyper spectral image processing, remote sensing, high definition television (HDTV), industrial X-ray image processing, microscopic imaging, other image/video processing applications and medical image processing [13].

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.

In the context of medical image, image enhancement also applied in many applications. Morphological transform was applied to medical images [19]. 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 11 shows the morphological transform was used in X-ray of renal system. Figure 11(a) presents the original image. Figure 11(b) illustrates optimum enhanced image. While Figure 11(c) and Figure 11(d) shows the transformed image with different radius using CNR and PSNR, respectively.



Figure 11 The example of X-ray image of renal system using morphological transform [19]

**2.2.3 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 [20]. With the respect to dimensional reduction there are some existing techniques including: Independent component analysis, Isomap, Kernel PCA, Latent semantic analysis, Partial least squares, Principal component analysis, 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. 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 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 method are broadly divided into three categories: (i) statistical methods, (ii) structural methods and (iii) model based methods [21]. A number of existing texture analysis models has been developed 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 [3] the long rage, 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 [22] described texture analysis that texture is an important approach for describing the region. The texture can be defined as three different categories: (i) Statistical such as smooth and coarse, (ii) Structural and (iii) Spectral.

In [12], Frequency Filtering texture analysis was applied using Fourier transform to calculate filter data as the equation (2)

VD᷈ = F \* VD ………………………………..(2)

where VD᷈ is the filter data from the convolution theorem [22] while F is the filter in the spatial domain.

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

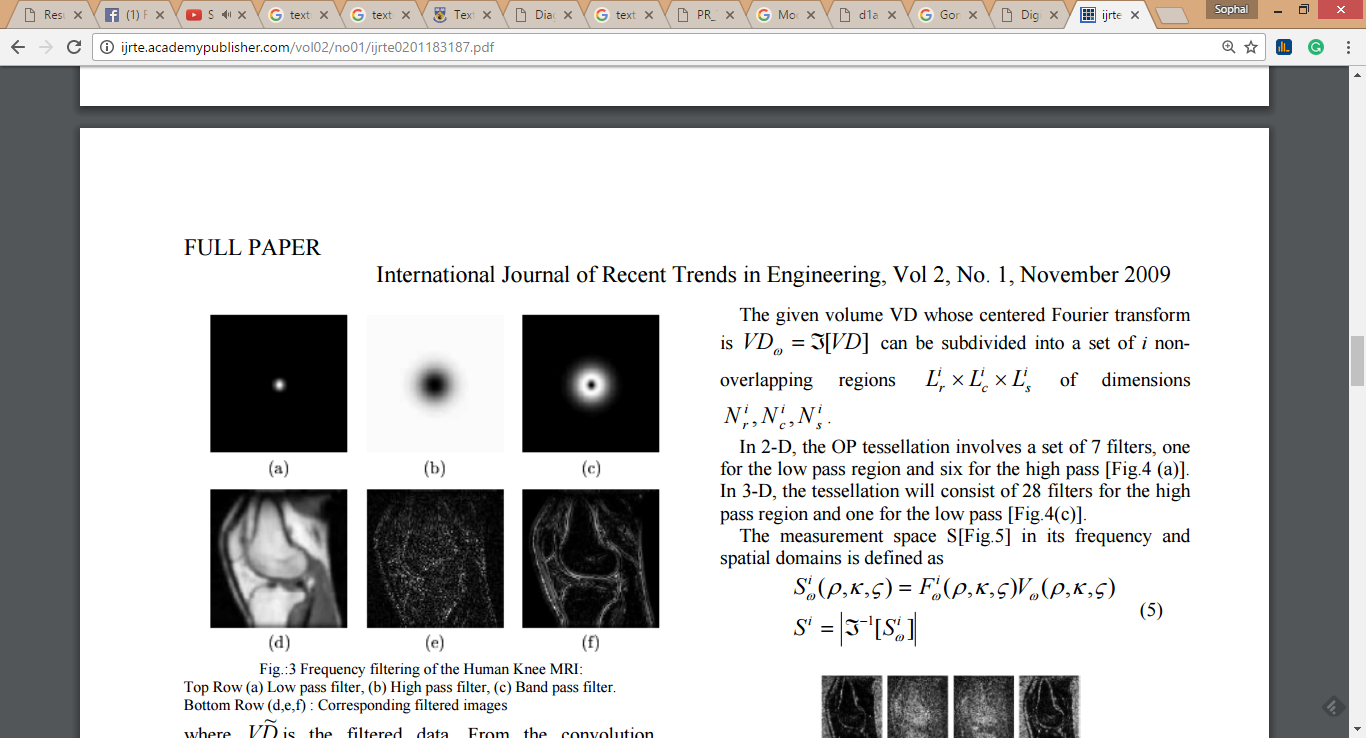


Figure 12 The example of various filters applied to MRI of human.

**Shape analysis**

Shape analysis is one of the branches in image science study which is directed with 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 [23]. 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 a number of applications. With respect to the work in this research is directed to Knee OA, the existing examples are found in [1, 11, 14].

In [1] the shape analysis was used to calculate the thickness between the bone and joint space in knee, the seven-steps process was proposed: Step 1: map the intensity values in greyscale image to new values in such a way that few data is saturated at low and high intensities of image. Thus the contrast of the output image was improved. Step 2: threshold the image with its mean intensity values. Step 3: edge detection of the image. Step 4: cropping the image. Step 5: detecting the boundary of joint space. Step 6: making the binary for cropped image. And the final step (Step 7) the distance calculation. As the result the 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.

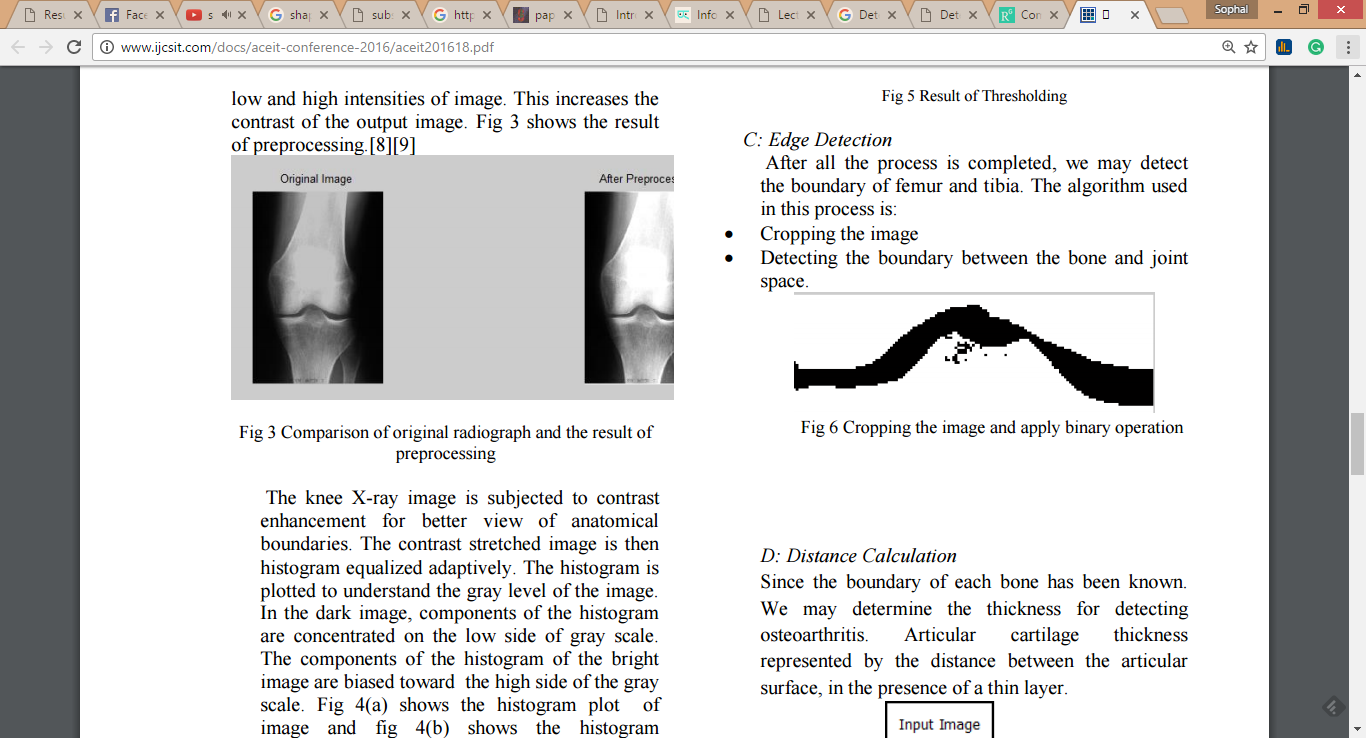


Figure 13 Cropping the image and apply in binary operation [1]

With the respect to the work presented in [14], 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 for calculating the area of cartilage by the number of pixel as presented in Figure 14.

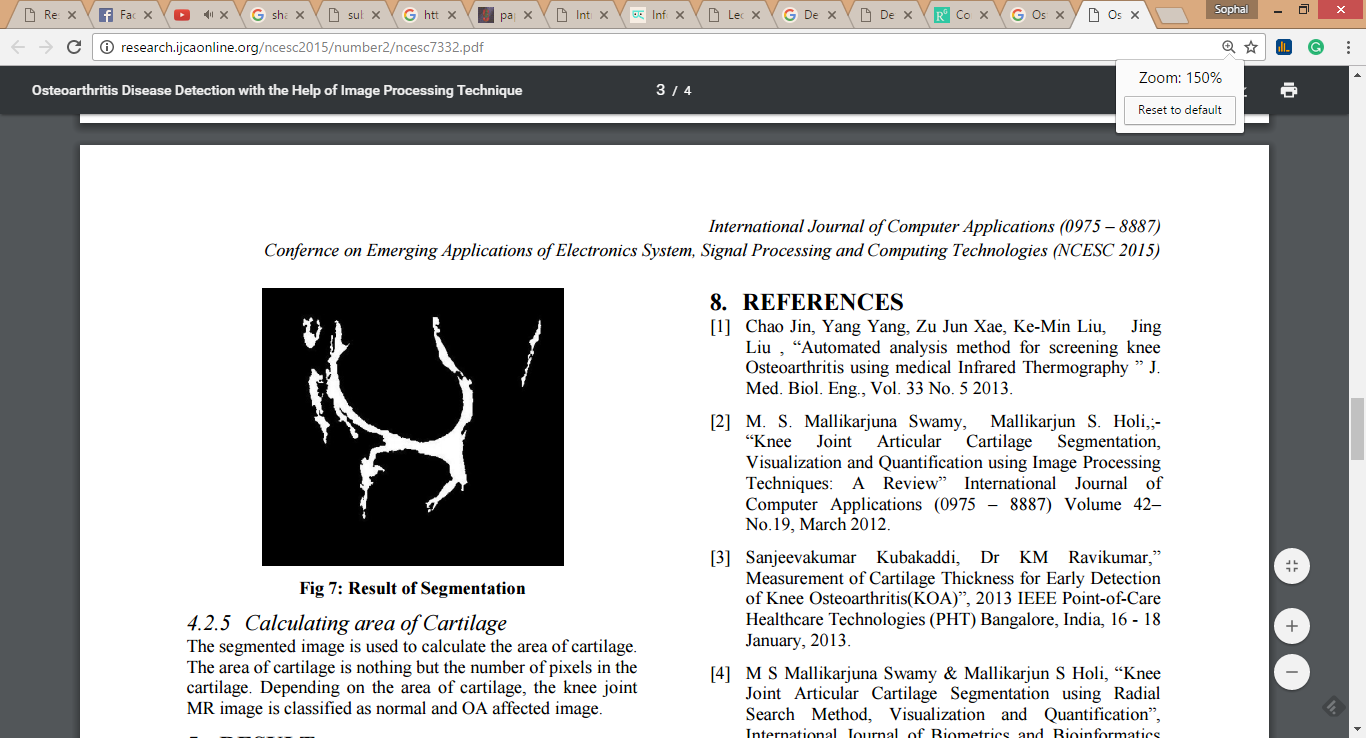


Figure 14 MRI image of knee after the texture filtering [14]

In summary, the image segementation, 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 segementation, enhancement and representation are applied as the pre-procesing process. The next section is 2.3 image classification is discussed.

**2.3 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. This section is organised as follows: the fundamental idea of data mining is presented in Sub-section 2.3.1. Sub-section 2.3.2 suggests the detail of image classification and finally Sub-section 2.3.3 presents the brief detail of classification learning methods.

**2.3.1 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 [24]. 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), finding presentation, finding interpretation and finding evaluation.

In [25] KDD consists of five processes: (i) Selection, (ii) Preprocessing, (iii) Transformation, (iv) Data mining and (v) Interpretation and evaluation as shown in Figure 15. From the Figure 15 can be observed that the KDD is a five-process mechanism. During selection process a large collection of data is selected. When selected data is delivered which only interested subset of data is choose the preprocessing process is then performed [25], an example of preprocessing process including the date attribute is computed into the day of week from a date field. After that the transformation process is then commenced in order to prepare appropriate data format to be ready for the data mining process. During the data mining process, the data mining algorithms or techniques are implemented. There existing data mining algorithms or techniques include: decision tree algorithms, logistic regression algorithms, and Naïve Bayes algorithm. When the pattern or hidden information is delivered the interpretation and evaluation process is applied to get the accurate knowledge.

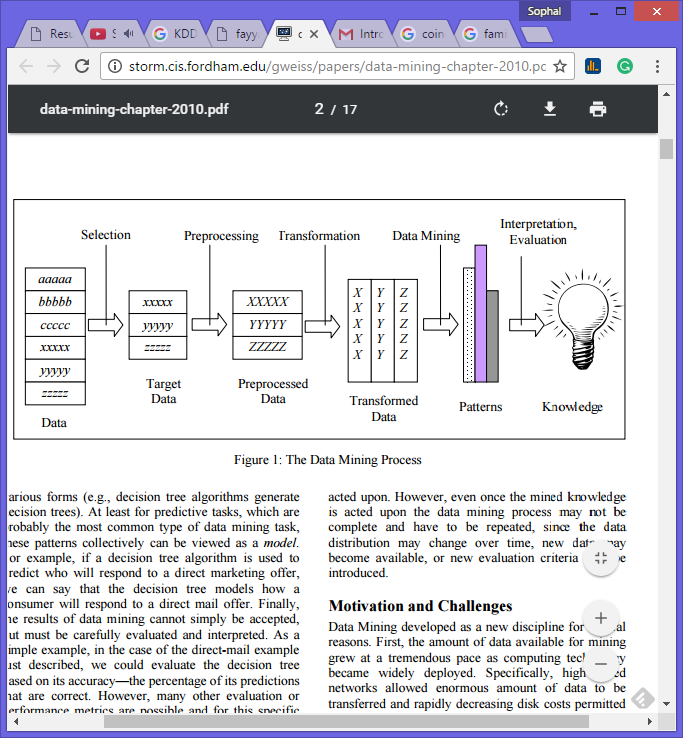


Figure 15 KDD process for producing the knowledge [15].

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.

Data mining can be applied to any kind of data repositories. In the traditional, data mining was applied to relational database (relational database is a database structure that can store data in tabular format which has two dimensional form row and column). In recently data mining can be applied to various of data formats such as spatial data, time-series data, free text, multimedia data, hypertext data and video and image data.

In [37] the task of data mining is generally divided into three main categories: (i) Classification, (ii) Association and (iii) Clustering.

Classification is the finding of a model that describes the data classes or concepts based on its attribute. The purpose of classification is to be able to use the derived model in order to predict the class of objects whose class label is unknown. Therefore, 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 together. Association rule reveals the associative relationship among the objects [37], for example: a Tesco in 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 [26]. Image mining consists of multiple components [27] such as: image analysis, image classification, image indexing, image retrieval and data management. The component of image mining is presented in Figure 16.

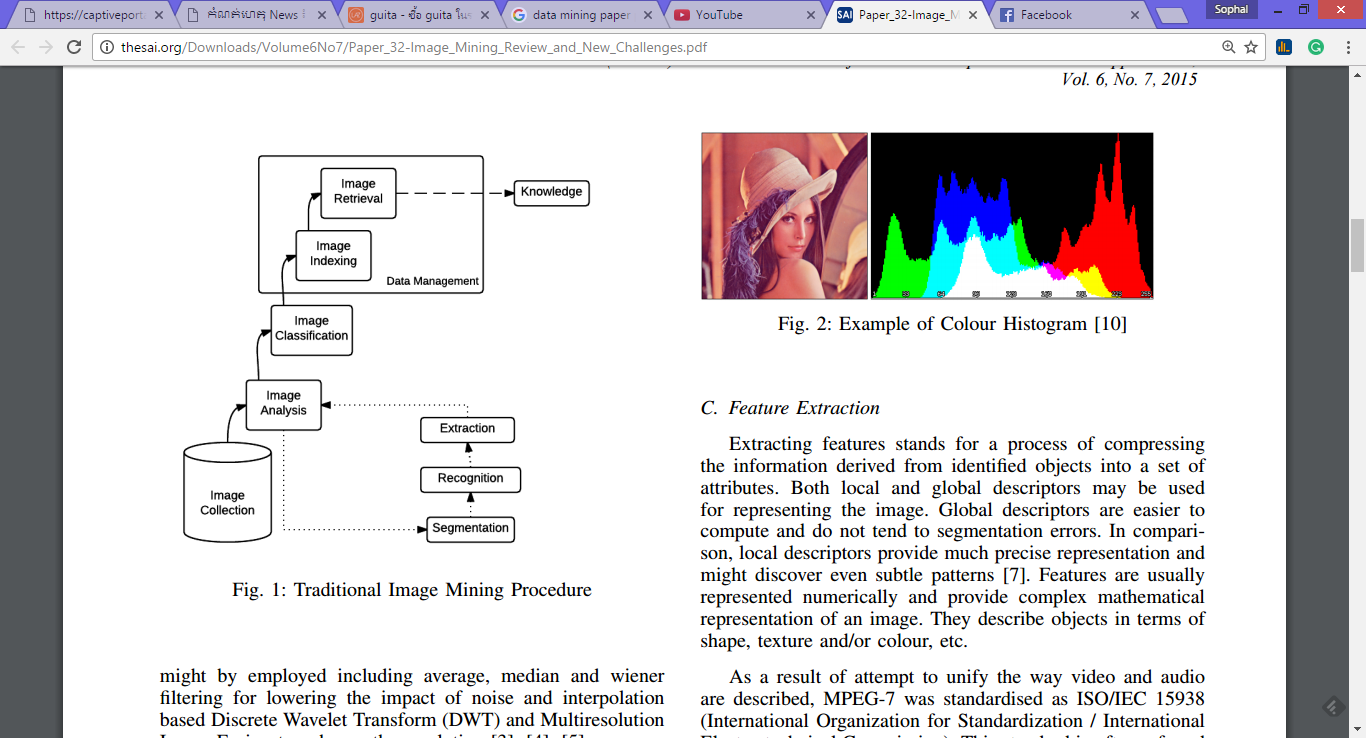


Figure 16 Traditional image mining process [27]

**2.3.2 Image Classification**

Image classification is a branch of image analysis, which 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 pixel property. Image classification is also widely used in medical image, statelier 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, (ii) spatial pattern recognition and (iii) temporal pattern recognition.

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.

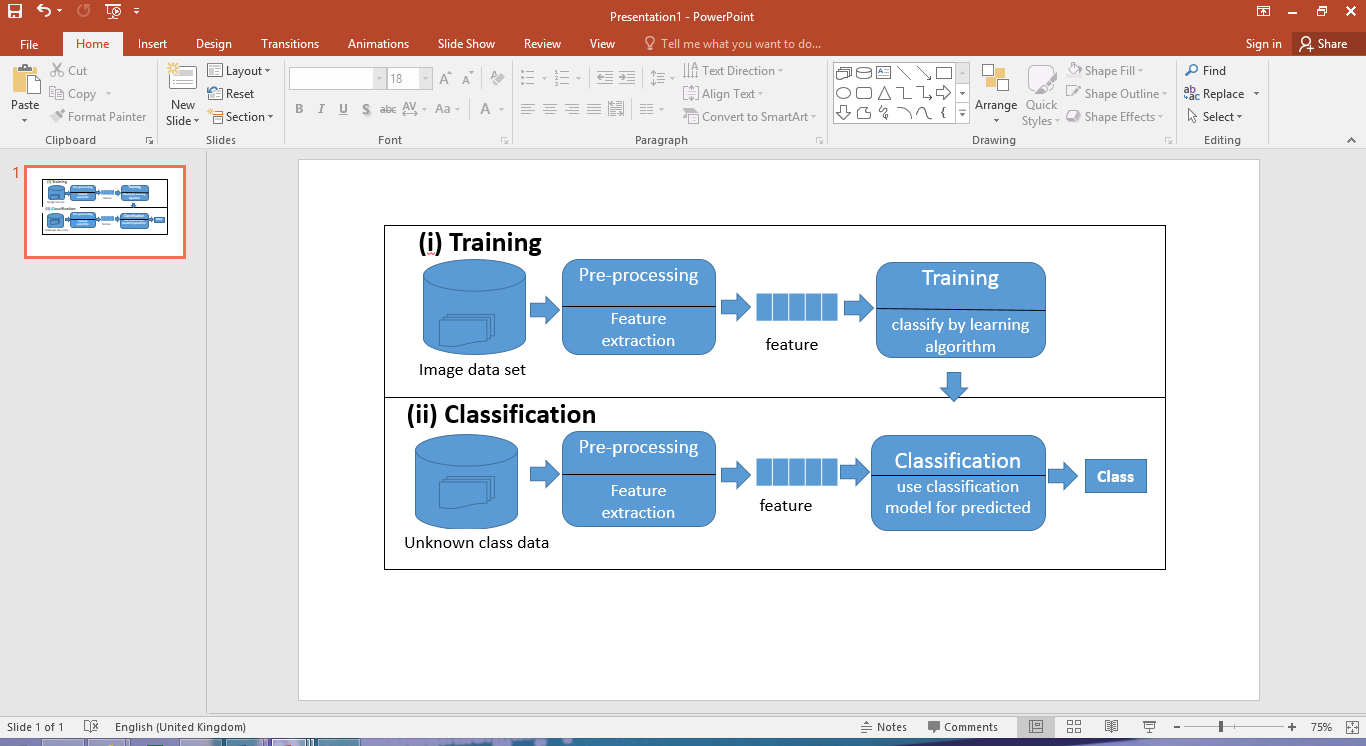


Figure 17 Image classification process

Image classification process is presented in Figure 17. From the Figure 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 {i1, i2, …, in-1, cn} where ii is an attribute value and cn is a class label that cn ∈ C.

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 the data class.

As noted above, image classification technique has two main process where in the classification application process need to predict on the usage of discrete class label. In order to make prediction concern with the real value, regression is suggested.

Regression is a statistical technique which used to analyse the relationship between a response variable and one or more predictor variable. Liner regression and multiple linear regressions are the frequently is the existing technique for regression analysis. Linear regression is the one of well-known regression which use to analyse the correlation between predictor variable (known as the independent, explanatory, regress, input or exogenous variable) considered as *x* and response variable (known as dependent, target, output or endogenous variable) denoted by y.

Simple linear regression (SLR) is the most well-known of linear regression where in SLR have only one predictor *x*. The equation 3 below presented the simple regressing model.

y = a0 + a1x ……………………(3)

where a0 and a1 are the constant for the regression. However, in real-life data the response variable (y) might have relevant incorrect related to each predictor variable (x) value. Residual value (*e)* is the differentiation of value between an actual response value and a predicted response value. The simple regressing model with residual value as shown in equation 4.

y = a0 + a1x + *e*……………………(4)

Multiple Linear Regression *(MLR*), a number of predictor variable is used. The equation 5 presents MRL model for predicting the response value.

y + *e* ………….…(5)

where y is the response variable, *x1, x2, ……., xn* are the predictor variables, *e* is a random error and a0, a1, ……, an are the regression coefficients extracted from input data.

Linear regression has the limitation of linear regression such as: (i) a lack of auto correlation and (ii) limited to numerical response prediction

* + 1. **Classification Learning Methods**

As mentioned before the work in this thesis is concerned to image classification. There are five classification learning methods are presented in this sub-section: (i) decision tree generator, (ii) Naïve Bayes and (iii) Logistic regression.

**Decision Tree**

Decision Tree is considered as the most popular and widely use 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 18 illustrates the decision tree for the direct mailing response.

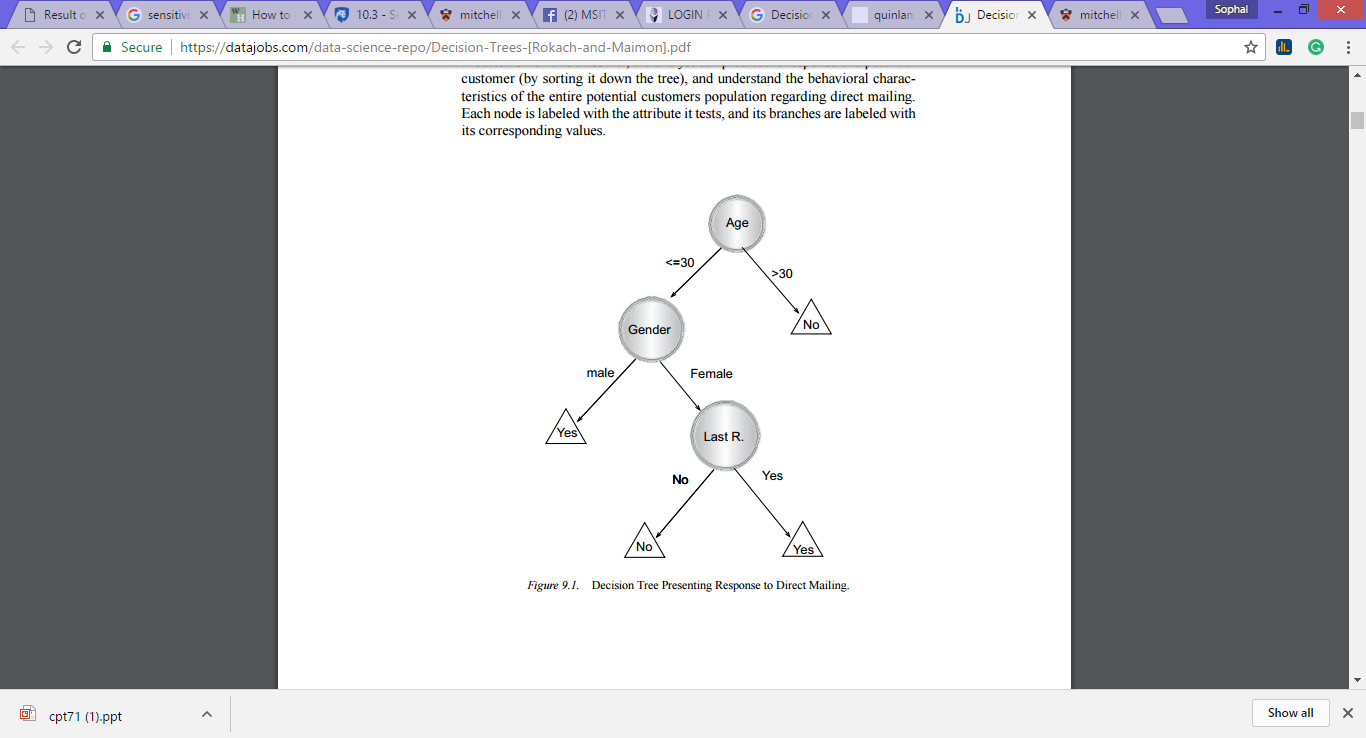


Figure 18 Response to Direct Mailing Decision Tree [38].

­**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 (6) below:

P = …………………………(6)

where is the probability of instance d being in class ci,  is probability of generating instance d given class ci,  is the probability of occurrence of class ci and p(d) = probability of instance d occurring.

**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 (7).

Logistic regression equation……………….(7)

where p is the probability of presence of the characteristic of interest.

**2.4 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 there are a number of evaluation measurement is applied: (i) Confusion Matrix, i(i) Accuracy, (iii) Recall, (iv) Sensitivity, (v) Specificity, (vi) F-Measure, (vii) Area Under Curve (AUC) and (viii) Comparison using Friedman test. The detail of each measurement method will be discussed below:

**2.4.1 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 presents the confusion matrix.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Predicted Class** | |
| True | False |
| **Actual**  **Class** | True | TP | TN |
| False | FP | FN |

Table 2 Confusion Matrix

From the Table 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) Rate error, (iii) Recall , (iv) Precision, (v) Sensitivity and (vi) Specificity. The equation for each measure is presented in the Equation 8 to Equation 13, respectively.

Accuracy = (TP +FN) / T, where T = TP + TN + FP + FN……………………(8)

Rate error = (TN+FN) / T……………………………………………………….….(9)

Recall = TP / (TP + TN) ………………………………………………………….(10)

Precision = TP / (FP + TP) ……………………………………………………….(11)

Sensitivity = [TP / (TP + FN) ] \* 100 ………………………………………..….(12)

Specificity = [TN / (TN + FP) ] \*100 ……………………………………..……..(13)

**2.4.2 Area Under Curve (AUC)**

Area under curve (AUC) is considered as the best method for indicating the overall quality of the classifier [27]. The AUC can be calculated from 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 [27]. The ROC graph is shown in Figure 18.

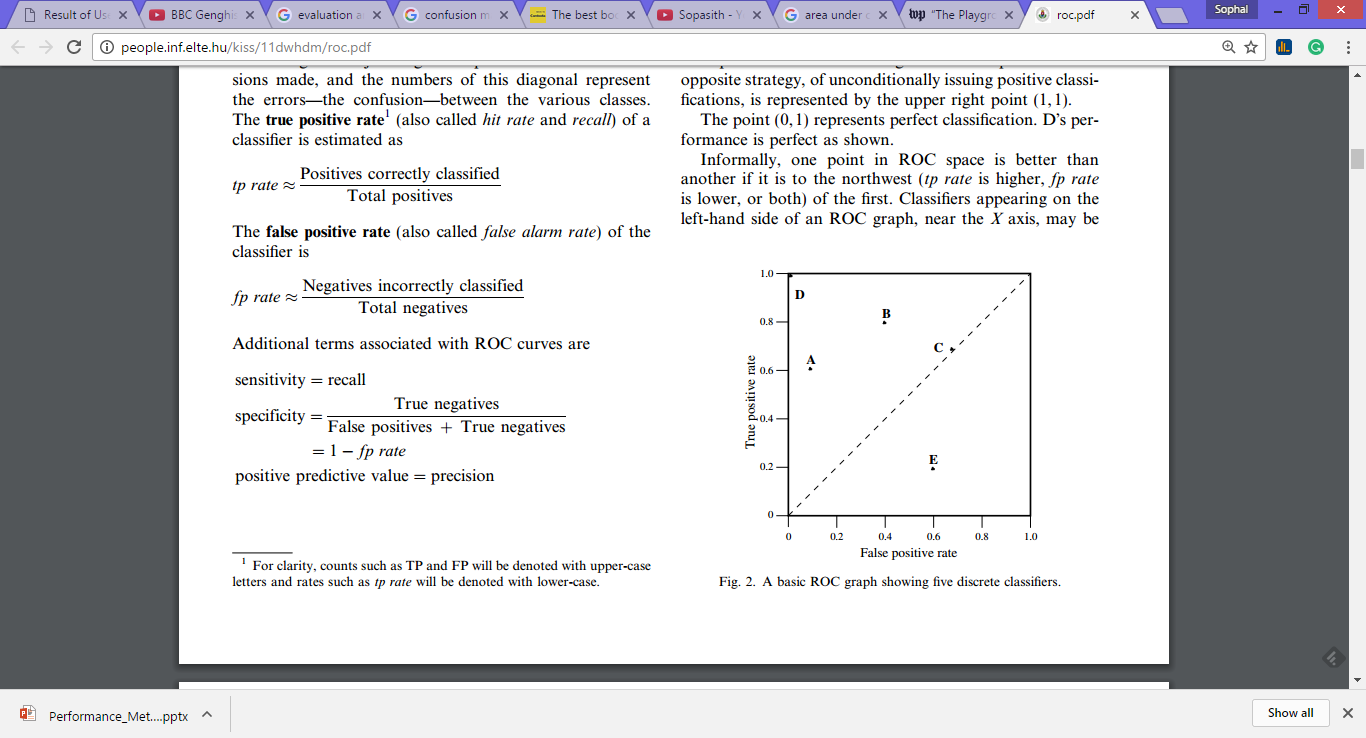


Figure 19 The ROC graph with 5 discrete classifiers [27]

AUC refers to the area under the ROC curve. Figure 19 presents the AUC in ROC graph which the random value from 0.0 to 1.0 in probabilistic classifier.

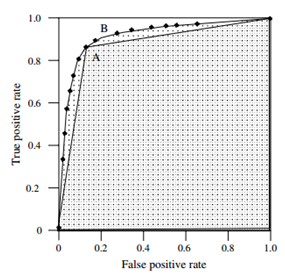
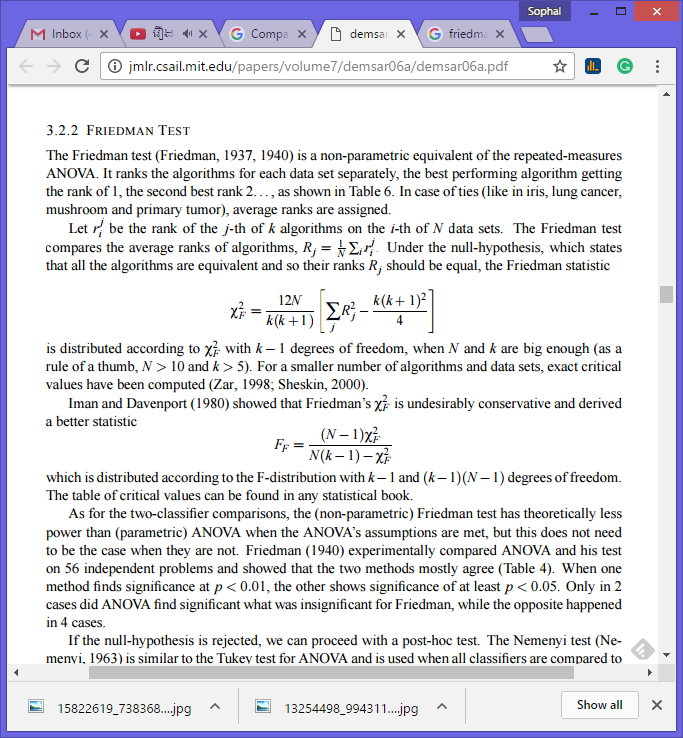


Figure 20 Area Under ROC Curve (AUC) [27]

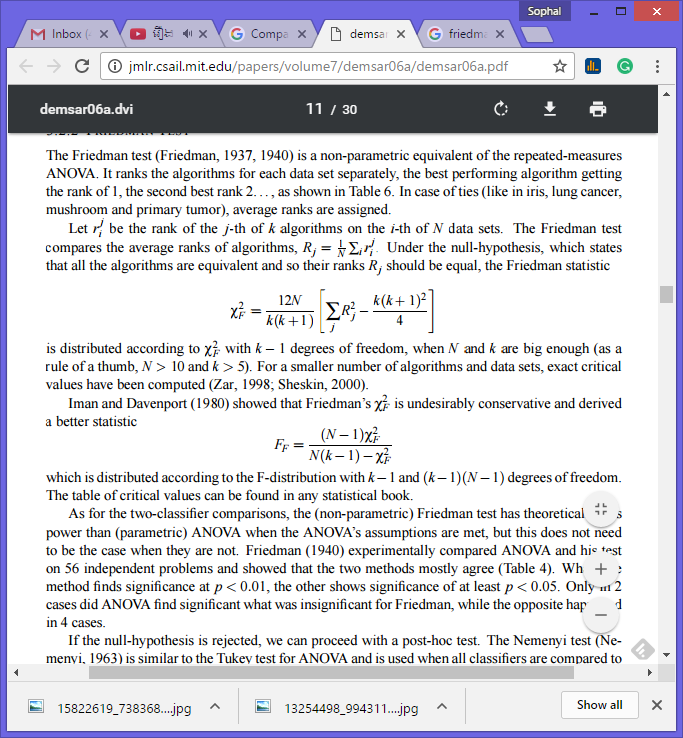
**2.4.3 Comparison Measure**

Friedman test is the statistical test used for classifier comparison of classifiers or across multiple data sets. Friedman first introduced the Friedman test, which is a non-parametric alternative of the repeated-measure ANOVA. The Friedman test is the test to show the rank of algorithm which algorithm is the best performance. For the best-performing algorithm get the first rank, the second best-performing algorithm get the second rank and so on.

The Friedman test can be calculated using the Equation 10. Let *rij* is the rank of the jth of k algorithms in the ith of N data sets. In this case, the implement of Friedman test for comparing the average ranks of algorithms, Rj =(1/N) , with null-hypothesis.

……………………………(14)

The χ 2 F with k − 1 degrees of freedom, when N and k are big enough (N>10 and k>5). Friedman’s χ 2 F is undesirably conservative and derived a better statistic that was introduced by Iman and Davenport (1980) as shown in Equation 11.

……………………….(15)

which is distributed according to the F-distribution with k−1 and (k−1)(N−1) degrees of freedom.

**2.5 Summary of the chapter**

Reference

[1] Mahima Shanker Pandey, S S Soam and Surya Prakash Tripathi, “Detection of Knee Osteoarthritis Using X-Ray”, ACEIT Conference, 2016

[2] Lior Shamir, Salim Rahimi, Nikita Orlov, Luigi Ferrucci and Ilya G. Goldberg, “Progression Analysis and Stage Discovery in Continuous Physiological Process Using Image Computing”, EURASIP Journal on Bioinformatics and Systems Biology, Vol 2010, 2010

[3] Shivanand S. Gornale, Pooja U. Patravali and Ramesh R. Manza, “Detection of Osteoarthritis using Knee X-Ray Image Analyses: A Machine Vision based Approach”, International Journal of Computer Application, Vol 145-No.1, July 2015.

[4] L. Shamir, S. M. Ling, W. Scott, M. Hochberg, L. Ferrucci and I. G. Goldberg, “Early detection of radiographic knee osteoarthritis using computer-aided analysis”, Sciencedirect , Elsevier Ltd on behalf of Osteoarthritis Research Society International, 2009.

[5] Lior Shamir, “A computer analysis method for correcting knee X-rays with continuous indicator”, Int J CARS (2011) 6:699-704, 2011.

[6] Lior Shamir, Shari M. Ling, William W. Scott, Jr., Angelo Bos, Nikita Orlov, Tomas J. Macura, D. Mark Eckly, Luigi Ferrucci and Ilya G. Goldberg, “Knee X-Ray Image Analysis Method for Automated Detection of Osteoarthritis”, IEEE Transaction on Biomedical Engineering, Vol. 56, No. 2, February 2009.

[7] Amir Herman, Ofir Chechik, Ganit Segal, Yona Kosashvili, Ran Lador, Moshe Salai, Amit Mor, Avi Elbaz and Amir Haim, “The correlation between radiographic knee OA and clinical symptoms-do we know everything?”, Clin Rheumatol (2015) 34:1955-1960, 2015.

[8] P. Podsiadlo, L. Dahl, M. Englund, L. S. Lohmander and G. W. Stachowiak, “Difference

in trabecula bone texture between knees with and without radiographic osteoarthritis

detected by fractal method”, Sciencedirect, Osteoarthritis and Cartilage, vol 16, 2007.

[9] J. H. Kellgren and J. S. Lawrence, “Radiologic assessment of osteoarthritis,” Annals of the Rheumatic Diseases, vol. 16, pp. 494–501, 1957.

[10] J. H. Kellgren, M. Jeffrey, and J. Ball, Atlas of Standard Radiographs, Blackwell Scientific, Oxford, UK, 1963.

[11] Sanjeevakumar Kubakaddi, Km Ravikumar and Harini DG, “Measurement of Cartilage Thickness for Early Detection of Knee Osteoarthritis (KOA)”, 2013 IEEE Point-of-Care Healthcare Technologies (PHT), India, January, 2013

[12] P.N. Jebarani Sargunar and Kalavakkam, “Texture Defect Detection Technique for MRI Knee Images”, International Journal of Recent Trend in Engineering, Vol 2, No. 1, November 2009.

[13] Xushu Zhang, Yuan Guo and Peng Du, “The Contour Detection and Extraction for Medical Images of Knee Joint”, Bioinformatics and Biomedical Engineering, (iCBBE) 2011 5th International Conference, 2011.

[14] Bhagyashri L. Wagaj and M.M. Patil, “Osteoarthritis Disease Detection with the Help of Image Processing Technique”, International Journal of Computer Application, Conference on Emerging Application of Electronic System, Signal Processing and Computing Technologies (NCESC 2015), 2015.

[15] Ch.Hima Bindu, “An Improved Medical Image Segmentation Algorithm Using Otsu Method”, International Journal of Recent Trends in Engineering, Vol 2, No. 3, November 2009.

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

[17] H.P. Ng, S. Huang, S.H. Ong, K.W.C Foong, P.S Goh and W.L. Nowinski, “Medical Image Segmentation Using Watershed Segmentation with Texture-Based Region Merging”, 30th Annual International IEEE EMBS Conference, August 2008.

[18] Samir . Kumar Bandyopadhyay, “Edge Detection in Brain Images”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2) , 2011.

[19] Raihan Firoz, Md. Shahjahan Ali, M. Nasir Uddin Khan, Md. Khalid Hossain, Md. Khairul Islam and Md. Shahinuzzaman, “Medical Image Enhancement Using Morphological Transformation.”, Journal of Data Analysis and Information Processing, 2016.

[20] Roweis, S. T.; Saul, L. K. (2000). "Nonlinear Dimensionality Reduction by Locally Linear Embedding". *Science*. **290** (5500): 2323–2326.

[21] Jianguo Zhang and Tieniu Tan. “Brief review of invariant texture analysis methods”. The journal of the pattern recognition society, Pattern Recognition 35 (2002) 735–747, 2002. #

[22] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins. “Digital Image Processing Using Matlab ”.(IND,2004).

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

[24] MOHAMMED J. ZAKI and WAGNER MEIRA JR. . “DATA MINING AND ANALYSIS Fundamental Concepts and Algorithms”. 2014.

[25] Gary M. Weiss and Brian D. Davison. “Handbook of Technology Management, H. Bidgoli (Ed.), John Wiley and Sons, 2010.”. 2010.

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

[27] Tom Fawcett. “An introduction to ROC analysis”. Science direct on Pattern Recognition Letter 27 (2006) 861–874. 2006.

[28] Janez Demsar. “Statistical Comparisons of Classifiers over Multiple Data Sets”. Journal of Machine Learning Research 7 (2006) 1–30. 2006.

[29] Ahmed Majid Taha, Aida Mustapha and Soong-Der Chen. “**Naive Bayes-Guided Bat Algorithm for Feature Selection**”. ScientificWorldJournal. 2013. 2013.

[30] Webb GI. Naïve bayes. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning. New York, NY, USA: Springer; 2010. pp. 713–714.

[31] A. Marcano-Cedeño, J. Quintanilla-Domínguez, M.G. Cortina-Januchs and D. Andina. “Feature Selection Using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network.”. IEEE. 2010.

[32] (retrieve from <http://www.the-scientist.com/?articles.view/articleNo/30693/title/The-First-X-ray--1895/>)

[33] retrieved from http://cw.tandf.co.uk/imagingforstudents/sample-material/Chapter-1-Introduction-to-Medical-Imaging.pdf)

[34](Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4639504>)

[35](retrieved from <https://www.spacetelescope.org/static/projects/fits_liberator/image_processing.pdf>)

[36] M. Jayathi, “Comparative Study of Different Techniques Used for medical image Segmentation of Liver from Abdominal CT Scan”, IEEE WiSPNET 2016 conference, 2016.

[37] Yongjian Fu. “Data Mining: Task, Techniques, and Application”.

[38] Lior Rokach and Oded Maimon. “DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK”. (Israel, 2010).