SEGMENTATION OF REGION OF INTEREST AND EXTRACTION OF SIGNIFICANT FEATURES FOR HEP-2 IMAGES

KHAW WIL BOND

UNIVERSITI SAINS MALAYSIA

2017

SEGMENTATION OF REGION OF INTEREST AND EXTRACTION OF SIGNIFICANT FEATURES FOR HEP-2 IMAGES

by

KHAW WIL BOND

A Dissertation submitted for partial fulfilment of the requirement
for the degree of Master of Science
(Electronic Systems Design Engineering)

August 2017

ACKNOWLEDGEMENT

Throughout the process of this research project, I have gained experience and acquired a lot of new knowledge that I have never exposed to. During the implementation of this research project, it is not only advanced in my technical and writing skills, but also has improved my communication skill. At here, I would like to express my sincere thankful for all people that have shared their precious knowledge and supported by helping me in completion of this dissertation.

First and foremost, I would like to acknowledge my deepest sincere gratitude and appreciation to Professor Dr. Nor Ashidi bin Mat Isa as my supervisor for the constant guidance, cooperation, warmest support, tolerance and advice along the period of completion of my research project, I am honoured to have a very expertise supervisor in the fields of Biomedical, Image Processing and Artificial Intelligent.

Secondly, my special thanks are given to all the coordinators and committee members of Master research project session 2016/2017, and the lecturers of the School of Electrical and Electronic Engineering who have sacrificed their times to share their knowledge and experience in research project implementation techniques, thesis and proceeding papers writing skills, and oral presentation skill for us.

This research work is supported by Research University Individual (RUI) grant by Universiti Sains Malaysia titled “Development of an Intelligent Auto-Immune Diseases Diagnostic System by Classification of HEp-2 Immunofluorescence Patterns”.

Last but not least, I wish to extend my heartiest gratitude to my beloved family and my fellow friends for their strong immense support, directly or indirectly, and absolute love, in this process of completion of this dissertation.

TABLE OF CONTENTS

ACKNOWLEDGEMENT……..... ii

TABLE OF CONTENTS……... iv

LIST OF TABLES……..... viii

LIST OF FIGURES……... ix

LIST OF ABBREVIATIONS…… xii

LIST OF SYMBOLS……..... xiv

ABSTRAK……..... xvi

ABSTRACT……... xviii

CHAPTER 1 – INTRODUCTION

1.1 Background…… 1

1.2 Problem Statement……..... 3

1.3 Research Objectives……... 4

1.4 Research Scope…….. 5

1.5 Dissertation Outline……... 5

CHAPTER 2 – LITERATURE REVIEW

2.1 Introduction…… 8

2.2 HEp-2 Cell……..... 9

2.3 Digital Image Processing……... 17

2.4 Segmentation of HEp-2 Cell……..... 21

2.4.1 Basic Concept of Segmentation in Image Processing……... 22

2.4.2 Segmentation Technique in HEp-2 Cell Image……..... 23

2.4.2.1 Thresholding…….. 24
2.4.2.2 Watershed…….. 26
2.4.2.3 K-Means Clustering……... 27
2.4.2.4 Fuzzy C-Means Clustering…… 28
2.4.3 Critical Analysis on State-of-the-art of Segmentation Techniques……..... 30
2.5 Feature Extraction…….. 32
2.5.1 Basic Concept of Feature Extraction in Image Processing…… 33
2.5.2 Feature Extraction Techniques in HEp-2 Cell Image…… 35
2.5.2.1 Gray Level Co-occurrence Matrix (GLCM)…….. 35
2.5.2.2 Normalized Histogram of Oriented Gradients (HOG)…….. 37
2.5.2.3 Speeded-Up Robust Features (SURF)……... 38
2.5.2.4 Local Binary Pattern (LBP)……... 40
2.5.3 Critical Analysis on State-of-the-art of Feature Extraction Techniques……..... 42
2.6 Summary……..…..… 44

CHAPTER 3 – METHODOLOGY

3.1 Introduction……..……..... 46
3.2 Simulation Tool and Data Samples……... 47
3.2.1 Simulation Tool……..... 47
3.2.2 Data Samples……..... 48
3.3 New Proposed Segmentation Technique……... 52
3.3.1 Overview of New Proposed Segmentation Technique…….. 52
3.3.2 Image Colour Map Conversion……..... 57
3.3.3 Image Pre-processing……..... 57

3.3.4 Segmentation by FCM Clustering……..... 58

3.3.5 Segmentation by Thresholding…….. 61

3.4 Feature Extraction and Analysis…… 62

3.4.1 Overview of Feature Extraction and Analysis……... 62

3.4.2 Feature Extraction with GLCM……..... 63

3.5 Evaluation Methods……... 68

3.5.1 Qualitative Analysis……... 69

3.5.2 Quantitative Analysis……..... 70

3.6 Summary…… 71

CHAPTER 4 – RESULTS AND DISCUSSIONS

4.1 Introduction……..... 73

4.2 Results for Segmentation of HEp-2 Images…….. 74

4.2.1 Segmentation Results for MIVIA Data Sample…… 74

4.2.2 Segmentation Results for HUSM Data Sample……..... 82

4.2.3 Discussion on Segmentation Results……..... 89

4.3 Results for Feature Extraction of HEp-2 Images……... 93

4.3.1 Feature Extraction Results for MIVIA Data Sample……..... 93

4.3.2 Feature Extraction Results for HUSM Data Sample……..... 96

4.3.3 Discussion on Feature Extraction Results……..... 99

4.4 Conclusion……..... 103

CHAPTER 5 – CONCLUSION AND FUTURE WORKS

5.1 Conclusion……..…..…. 104

5.2 Future Works……..... 105

REFERENCES……..... 107

LIST OF TABLES

Table 2.1	Six common staining patterns of HEp-2 cells	13
Table 2.2	Characteristics of each class of staining patterns of HEp-2 cells	15
Table 2.3	Summary on the advantages and limitations of the reviewed segmentation technique	31
Table 2.4	Summary on the advantages and limitations of the reviewed feature extraction technique	43
Table 3.1	MIVIA and HUSM HEp-2 image data sample	52
Table 3.2	Properties of features to be extracted	67
Table 4.1	Summary of relationship between extracted features and staining pattern of HEp-2 cells for MIVIA	100
Table 4.2	Summary of relationship between extracted features and staining pattern of HEp-2 cells for HUSM	101

LIST OF FIGURES

Figure 2.1	Example of HEp-2 cell image	10
Figure 2.2	Summary of the two main steps in manual IIF test where (a) Step 1: Observation of Cells and (b) Step 2: Classification	12
Figure 2.3	Block diagramof the steps in digital image processing	21
Figure 3.1	Data sample for Fine Speckled pattern from (a) MIVIA and (b) HUSM	51
Figure 3.2	Flowchart of proposed segmentation algorithm for HEp-2 cell images	55
Figure 3.3	Segmented HEp-2 cell image with 3 clusters (black, gray and white)	60
Figure 3.4	Block diagram of feature extraction process by GLCM	63
Figure 3.5	Example of how GLCM works where (a) pre-process of input image before GLCM and (b) calculation of GLCM	65
Figure 3.6	Sample box-and-whisker plot	71
Figure 4.1	Segmented images for MIVIA Centromere. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm.	76
Figure 4.2	Segmented images for MIVIA Nucleolar. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm.	77

Figure 4.3 Segmented images for MIVIA Homogeneous. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 78

Figure 4.4 Segmented images for MIVIA Cytoplasmic. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 79

Figure 4.5 Segmented images for MIVIA Fine Speckled. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 80

Figure 4.6 Segmented images for MIVIA Coarse Speckled. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 81

Figure 4.7 Segmented images for HUSM Centromere. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 83

Figure 4.8 Segmented images for HUSM Nucleolar. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 84

Figure 4.9 Segmented images for HUSM Homogeneous. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm. 85

Figure 4.10	Segmented images for HUSM Cytoplasmic. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm.	86
Figure 4.11	Segmented images for HUSM Fine Speckled. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm.	87
Figure 4.12	Segmented images for HUSM Coarse Speckled. (a) Original HEp-2 image. (b) Segmented by Thresholding. (c) Segmented by FCM. (d) Segmented by K-Means Clustering. (e) Segmented by Watershed. (f) Segmented by proposed algorithm.	88
Figure 4.13	Contrast feature for MIVIA data sample	94
Figure 4.14	Correlation feature for MIVIA data sample	94
Figure 4.15	Energy feature for MIVIA data sample	95
Figure 4.16	Homogeneity feature for MIVIA data sample	95
Figure 4.17	Entropy feature for MIVIA data sample	96
Figure 4.18	Contrast feature for HUSM data sample	97
Figure 4.19	Correlation feature for HUSM data sample	97
Figure 4.20	Energy feature for HUSM data sample	98
Figure 4.21	Homogeneity feature for HUSM data sample	98
Figure 4.22	Entropy feature for HUSM data sample	99

LIST OF ABBREVIATIONS

ANA	AntiNuclear Autoantibody
B	Blue
bmp	bitmap
CAD	Computer Aided Diagnosis
CCD	Charge Coupled Device
CE	Centromere
CMY	Cyan, Magenta, and Yellow
CoALBP	Co-occurrence among Adjacent Local Binary Pattern
CS	Coarse Speckled
CT	Computed Tomography
CY	Cytoplasmic
DNA	DeoxyriboNucleic Acid
FCM	Fuzzy C-Means
FS	Fine Speckled
G	Green
GLCM	Gray Level Co-occurrence Matrix
HEp-2	Human Epithelial type 2
HO	Homogeneous
HOG	Histogram of Oriented Gradients
HUSM	Hospital Universiti Sains Malaysia
ICPR	International Conference on Pattern Recognition
IIF	Indirect ImmunoFluorescence

LBP	Local Binary Pattern
MATLAB	MATrix LABoratory
MD-LBP	MultiDimensional Local Binary Pattern
MRI	Magnetic Resonance Imaging
NaN	Not a Number
NU	Nucleolar
R	Red
RGB	Red, Green, and Blue
ROI	Region of Interest
RSURF	Rotation Speeded-Up Robust Features
SURF	Speeded-Up Robust Features
tif	tagged image format

LIST OF SYMBOLS

C	number of clusters in Fuzzy C-Means clustering
D	dimensional
$f(x, y)$	a two-dimensional function / an input image
$g(x, y)$	a thresholded image
$I_{grayscale}$	image in grayscale
i	a data point / grayscale intensity value / targeted pixel in an image
(i, j)	element in GLCM matrix
J_m	result of Fuzzy C-Means clustering
j	a cluster point / next targeted pixel of i
K	number of clusters in K-means clustering
M	number of rows of the digital image
M	fuzzy partition matrix exponent
N	number of columns of the digital image
N_g	number of grayscale
n	number of samples
$p(i, j)$	sum of number of times that pair of i and j occurs
r	radius
T	threshold value
μ_{ij}	the degree of membership of v_i in the j-th cluster
μ_j	the center of j-th cluster
v_i	the i-th measured data point
x	horizontal axis of continuous spatial coordinate

(x, y)	coordinate on the spatial domain
y	vertical axis of continuous spatial coordinate

SEGMENTASI KAWASAN YANG DIKEHENDAKI DAN PENGEKSTRAKAN CIRI-CIRI PENTING UNTUK IMEJ HEP-2

ABSTRAK

Imej epsitelium manusia jenis kedua (HEp-2) sangat penting dalam pengesanan *antinuclear autoantibody* (ANA) semasa menjalani diagnosis terhadap penyakit autoimun dalam badan manusia. Umumnya, sel HEp-2 boleh dibahagi kepada enam jenis iaitu Centromere, Nucleolar, Homogeneous, Cytoplasmic, Fine Speckled and Coarse Speckled. Walau bagaimanapun, dalam teknologi semasa, imej HEp-2 hanya boleh dianalisa secara manual dalam ujian immunofluorescence tidak langsung (IIF). Keputusan daripada ujian IIF menunjukkan kebolehubahan yang tinggi dan sangat bergantung kepada pengalaman ahli-ahli pakar fizik. Oleh itu, penyelidikan untuk mengubahsuai ujian IIF secara berdigital telah menarik minat para penyelidik termasuk dalam penyelidikan ini. Segmentasi dan pengekstrakan ciri-ciri daripada imej HEp-2 akan difokus dalam penyelidikan ini. Dalam segmentasi imej HEp-2, kaedah yang sedia ada gagal menghasilkan keputusan yang memuaskan. Oleh itu, satu kaedah baharu yang terdiri daripada gabungan dua kaedah konvensional iaitu *Fuzzy C-Means* (FCM) dan *thresholding* telah dicadangkan. Keputusan menunjukkan imej yang disegmentasi adalah lebih licin, konsisten dan mengandungi hingar yang rendah berbanding dengan keadah lain yang sedia ada. Dalam bahagian pengekstrakan ciri-ciri, kajian ini mengekstrak lima ciri iaitu *Contrast, Energy, Correlation, Homogeneity,* dan *Entropy*. Dari keputusan yang diperolehi, lima ciri yang dicadangkan berjaya membezakan corak-corak sel HEp-2.

Kesimpulannya, kaedah yang dicadangkan dalam penyelidikan ini menpunyai keupayaan yang tinggi untuk diperkenalkan dalam hospital untuk mengesan penyakit autoimun. Kaedah yang dicadangkan menpunyai ketepatan yang lebih tinggi dan boleh mengurangkan kelemahan yang terdapat dalam ujian IIF yang sedia ada.

SEGMENTATION OF REGION OF INTEREST AND EXTRACTION OF SIGNIFICANT FEATURES FOR HEP-2 IMAGES

ABSTRACT

Human Epithelial type 2 (HEp-2) images are important in detecting the antinuclear autoantibody (ANA) in diagnosis of autoimmune disease in human body. Generally, HEp-2 cells can be classified into six main patterns, namely Centromere, Nucleolar, Homogeneous, Cytoplasmic, Fine Speckled and Coarse Speckled. However, in current technology, HEp-2 images can only be analysed manually by indirect immunofluorescence (IIF) test. The result of IIF test has very high variability and very dependent on the experience of physicists. Therefore, digitalize the IIF test becomes the new interest to researchers as well as in this research, where segmentation and features extraction of HEp-2 images will be focused. In segmentation of HEp-2 images, the current state-of-the-art techniques failed to provide a satisfied segmented result. Therefore, a combination of two conventional methods (i.e. Fuzzy C-Means (FCM) clustering and thresholding) has been proposed in this study. From the result, the segmented images are smoother, more consistent and with lesser noises compared to other state-of-the-art methods. In feature extraction stage, this study proposes to extract five features, which are Contrast, Energy, Correlation, Homogeneity, and Entropy. Based on the results obtained, the five proposed features can successfully differentiate the staining patterns of HEp-2 cells. In short, the proposed methods in this research have high capability to be introduced in hospital for detection of HEp-2 images for

autoimmune disease. The proposed method has been proven with higher accuracy which can reduce the shortcoming of the existing IIF test.

CHAPTER 1

INTRODUCTION

1.1 Background

A picture or an image is a very powerful source that contains a lot of useful information, such as text, colour, shape, texture and size of objects. An English idiom "A picture is worth a thousand words" is best describe the powerful of an image. Because of this, human always make use of images in different applications, medical image processing is one of them. With the advancement of technologies, scientists and physicists have developed many new imaging inspection technologies to prevent and diagnose the new sickness. These new inspection techniques have a great effect in helping doctors to identify the sickness more accurately and able to take any further action in a faster time [1].

One of the sickness that becomes the interest in the medical image processing field is the autoimmune disease. An autoimmune disease is an illness that happens when the immune system in human body abnormally attacks the healthy tissues. Autoimmune disease can cause fatal because it may cause the healthy tissues and organ inside the patient's body to be destroyed [2]. Currently, there is a method used to detect the autoimmune disease in human body, known as indirect immunofluorescence (IIF) test. IIF test can detect the antinuclear autoantibody (ANA) in human epithelial type 2 (HEp-2) cells with the help of a fluorescence microscope.

ANA is an abnormal antibody found in human body that causes the autoimmune disease. Therefore, in order to detect and cure the mentioned disease, detecting ANA is the method. HEp-2 cells play a very important role in detecting the ANA because by observing the HEp-2 cells, physicists can determine whether the human body tissue is normal or infected. However, in the current technology, the processes of detecting the ANA from HEp-2 cells are still being carried out manually. This causes problem such as longer time consuming and inaccurate tested result because it fully depends on the experience of the physicists.

In current years, researchers started to develop a new automated algorithm to assist the physicists in detecting the autoimmune disease by analysing the HEp-2 cell images. The developed algorithm includes a few image processing steps such as segmentation and feature extraction based on HEp-2 cell images. Many existing known techniques have been studied and further developed to more accurately segment the HEp-2 cell from the original images. However, the results are still having rooms for improvement. Based on the segmented images also, the researchers have studied the most suitable features to be extracted that will be used for classification of the HEp-2 patterns. Because of the variability of the segmented result, no feature can be concluded as the best feature to classify the HEp-2 patterns. Therefore, a new improved version of segmentation algorithm is needed to improve the result and to determine the most suitable features to be extracted from the HEp-2 cell images.

1.2 Problem Statement

IIF is a great technique in detecting the ANA. However, this IIF technique suffers from a large variability coming from the lack of quantitative information, the photo bleaching effect, the varieties of reading systems and optics, the low standardization of the method and the most important, the ANA test depends much on the experience and expertise of the physicians [3]. Because of this variability, the reliability of the result from the existing ANA test is very low. On the other hand, the existing ANA test is very time consuming because it is not automated. With the current method of testing, physicians need to manually inspect the slides with the help of a fluorescent microscope and manually interpret the patterns slide by slide. Therefore, an automated algorithm for ANA test is needed to improve the weaknesses of existing ANA test.

From the past research, most of the segmentation algorithms have involved the thresholding method. Thresholding is the most frequently used method in application involving segmentation of images. However, in segmenting the HEp-2 cell images, some of the local information may be missing in the segmented images due to the inconsistent contrast and image quality of HEp-2 cell images. The pre-set threshold value needed for the segmentation have to be always adjusted to suit for different images. Other segmentation algorithm such as clustering related segmentation method is also popular in developing the algorithm to segment the HEp-2 cell images. However, this method is very sensitive to the noises in the images and the segmented result is not always consistent.

There are many features can be extracted from an image, such as entropy, contrast, energy, correlation, homogeneity, variance, skewness and kurtosis. However, in analysing the features extracted from the segmented HEp-2 images, not all of the

mentioned features are suitable to be used. Some of the extracted features have no significant different between images with different patterns. Besides, the feature of an image to be extracted is very depending on the quality of the image itself. If the quality of the segmented HEp-2 images is not good, for example, many noises can be found in the image, the extracted features will become inaccurate. Inaccurate features will affect the result of classifying the HEp-2 cell patterns. Therefore, it is important to extract only the suitable and accurate features from the segmented HEp-2 images.

In short, in order to overcome the aforementioned limitations in the current segmentation algorithm of HEp-2 cell images, a new improved segmentation algorithm must be developed as research motivation. The segmented results from the algorithm must be able to accurately differentiate the HEp-2 cell from the background and must be always consistent. Also, experiment has to be carried out on the segmented HEp-2 images to determine the most suitable features to be extracted in order to improve the accuracy in classification of HEp-2 cell pattern.

1.3 Research Objectives

The objectives of this research project are as follows:

i. To formulate a new improved segmentation algorithm from a combination of Fuzzy C-Means (FCM) clustering and thresholding for HEp-2 cell images.
ii. To analyse the extracted features from the segmented HEp-2 cell images and determine the most suitable features that can be used for classification of HEp-2 cell pattern based on Gray Level Co-occurrence Matrix (GLCM) technique.

1.4 Research Scope

In order to achieve the objectives, this research project will focus on two main parts, which are segmentation and feature extraction of HEp-2 cell images. The research project will start with development of an improved segmentation algorithm based on existing techniques which are thresholding and Fuzzy C-Means (FCM) clustering algorithm. The medical image implemented in this study is HEp-2 cell images from MIVIA and Hospital Universiti Sains Malaysia (HUSM). In this research project, the HEp-2 cell images will be processed in grayscale only.

The second part of this research project is focusing on feature extraction of HEp-2 cell images. Gray Level Co-occurrence Matrix (GLCM) is the selected technique used to extract the features from the images. The target images for this part of research will be the segmented HEp-2 cell images from the proposed segmentation algorithm developed in the first part of the research. The features to be studied and extracted in this research are Contrast, Energy, Correlation, Homogeneity, and Entropy. The development and simulation of the proposed segmentation algorithm as well as the study of feature extraction are achieved using MATrix LABoratory (MATLAB) R2015b.

1.5 Dissertation Outline

This dissertation is comprised of five main chapters which organize the entire study of this research project. In Chapter 1 (Introduction), the topic of the research with the motivation for requiring the improved segmentation algorithm for HEp-2 cell images as well as the study on the feature extraction based on the segmented HEp-2 images will

be briefly described. This chapter also contains problem statement, research objectives, research scope and dissertation outline of this research project.

In Chapter 2 (Literature Review), several basic literature reviews on the research area related to this dissertation are provided. These include an overview of the background and characteristic of HEp-2 cell as well as the fundamental and steps in digital image processing. Besides, this chapter also reviews the existing known segmentation techniques as well as feature extraction techniques on HEp-2 cell images. Comparisons on the advantages and limitations between the several segmentation techniques as well as the feature extraction techniques are stated in this chapter. Features that can be extracted based on the existing feature extraction techniques are also stated in this chapter.

In Chapter 3 (Methodology), the simulation tools and data samples to be implemented in this research are introduced. Besides, the process steps of each major development phases in both segmentation and feature extraction processes are also explained in detail. The evaluation methods for the proposed algorithm will be discussed atthe end of this chapter.

In Chapter 4 (Results and Discussion), the simulated experimental results of proposed segmentation algorithm and other state-of-the-art methods are presented for comparison and discussion. The performance and effectiveness of the proposed technique and previous approaches are evaluated using qualitative analysis. On the other hand, the simulated experimental results for extracted features based on selected technique are also presented and discussed. The effectiveness of the results is evaluated using quantitative analysis.

In Chapter 5 (Conclusion), the findings of the overall research works in this dissertation will be finalized. The significances and contributions of this research work will also be highlighted. In addition, future improvements are suggested for this developed system in order to improve the overall performance of the proposed algorithm in this research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Digital image processing is no longer a strange term to this world, especially in these recent decades. Nowadays, everything is digitalized, for example, from holding a book on hand to study to reading an e-book on digital devices or internet. Humans are preferred with digital system and application because of convenience. Similar to medical field, as the development of the world is getting advance, a lot of new sickness and diseases have been discovered. Many new imaging inspection technologies have been developed to prevent and to diagnose the new sickness. Because of this, hospitals have collected massive medical images. Therefore, if there is a method to analyse the medical images with a great significance effect, this will certainly help the doctors to detect and diagnose the disease [1]. Digital image processing is the method that can perfectly analyse the medical images, and therefore, it has become one of the popular research fields.

In digital image processing, especially for medical image processing, the characteristic or details of the images are very important on analysis and classification of diseases. Because of this, steps such as segmentation of images as well as feature extraction steps in extracting the important characteristics should be paid attention. Before these steps are carried out, it is important to understand and define which characteristic or features of the medical images should be focused. In this study, the image of HEp-2 cells will be studied.

This chapter will discuss on the existing works done by other researchers. The common HEp-2 cell patterns will first to study. Generally, the HEp-2 cell patterns consist of Centromere, Nucleolar, Homogeneous, Cytoplasmic, Fine Speckled and Coarse Speckled. Next, the concept of digital image processing will be studied to understand the important steps in digital image processing. In Section 2.4, methods of segmentation proposed by other research work will be reviewed. The techniques used in feature extraction will also be reviewed in Section 2.5. Lastly, this chapter is concluded in Section 2.6.

2.2 HEp-2 Cell

An autoimmune disease is an illness that happens in human body when the immune system is having inappropriate autoimmune response towards the healthy tissues in human body. Patients with autoimmune diseases always have unusual antibodies circulating in their blood which without noticed by the patients themselves. The unusual antibodies are produced by malfunction immune system in the patient's body and the unusual antibodies will attack the patient's own body tissues. There are more than eighty known autoimmune diseases and some of the well-known autoimmune diseases are multiple sclerosis, rheumatoid arthritis and diabetes mellitus type I [2]. The autoimmune disease is considered as a fatal disease because it will continuously destroy the healthy tissues and organs inside the human body. If no immediate action is taken, the chance for other simultaneous autoimmune disease to develop in human body may increase. Therefore, the ability to detect the autoimmune disease before it is too late is a big challenge to the physicians.

In order to detect and diagnose the autoimmune disease in human body, physicians must be able to detect the antinuclear autoantibody (ANA) in human epithelial type 2 (HEp-2) cells. HEp-2 cell is a kind of substrate that contains hundreds of autoantigens. Figure 2.1 shows an example of the HEp-2 cell image. From the figure, the green colour with oval shape represents the HEp-2 cells and the section in black colour is the background of the image. The cell membrane is the wall surrounding the HEp-2 cell and cytoplasm refers to gel-like substance enclosed within the cell membrane. When detecting the presence of ANA, HEp-2 cell will be mixed with the sample of human tissue [4]. ANA is a general term where it includes different type of autoantibodies found in human body that react with constituents of cell nuclei such as DeoxyriboNucleic Acid (DNA), proteins and ribonucleoproteins. Generally, ANA has a great function in determination of the presence of connective tissue [3, 4]. Detection of ANA is important because positive detection of ANA in patient's body tissue means the patient is suffering from autoimmune disease.

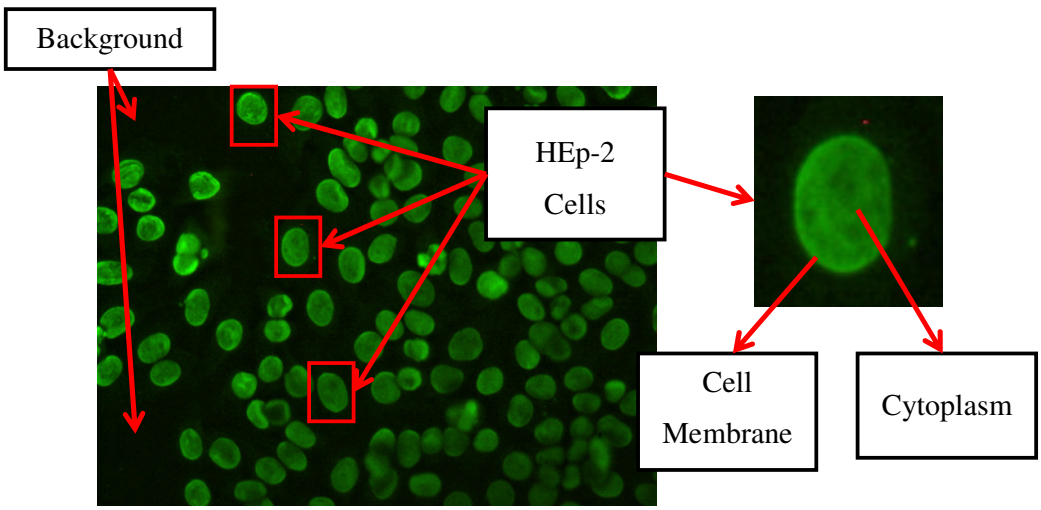
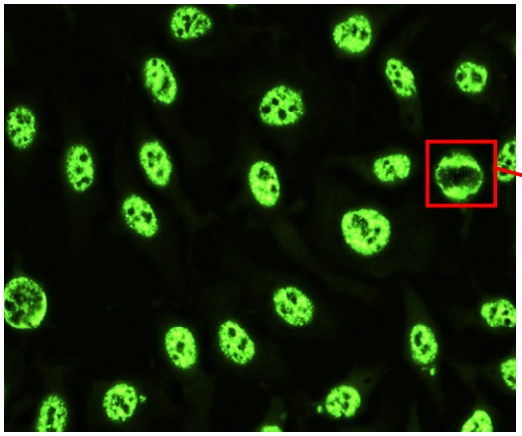


Figure 2.1: Example of HEp-2 cell image

Currently, the most popular method to carry out the detection of ANA is known as indirect immunofluorescence (IIF) test. IIF is a technique used to analyse microbiological samples with the help of a fluorescence microscope [2, 5]. By using the IIF technique together with HEp-2 cell, a broad range of autoantibody from the samples taken from patients will be scanned. Generally, the manual IIF test can be summarized into two main steps, which are observation of cells and classification. In the first step of IIF test, with the help of a fluorescence microscope, physicians will observe the fluorescence intensity, the numbers of mitotic cells and the patterns of the HEp-2 cells [2, 3, 6].

The next step of IIF test will be classifying the three mentioned criteria into different groups or patterns. During the classification stage, physicians will first categorize the HEp-2 cell samples into three different levels, namely positive, intermediate, and negative, based on the fluorescence intensity during the observation [2, 6]. For the category of positive and intermediate, physicians will further classify the samples into different patterns based on the presence of mitotic cells. If the number of observed mitotic cells is below a certain threshold, which is normally '1' or '2', the cells will be classified as negative or not infected [6]. Those results that are not eliminated will be analysed by physicians on the staining patterns of the cells. The summary of the two main steps in manual IIF test is shown in Figure 2.2. It is important for the physicians to be able to identify the patterns correctly to perform the appropriate follow-up action to confirm the diagnosis [5].

Step 1: Observation of Cells



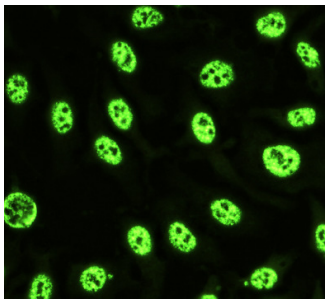
1. **Fluorescence Intensity**
   - Observation on the colour of the HEp-2 cells

2. **Number of Mitotic Cell**
   - Total 26 mitotic cells are observed

3. **Pattern**
   - To be determined based on six common patterns

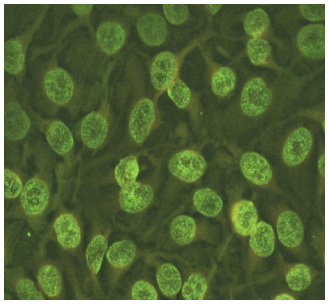
(a)

Step 2: Classification



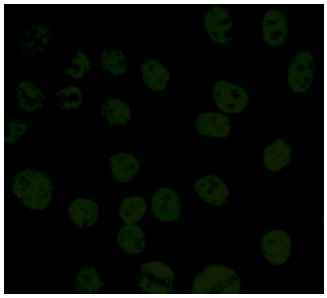
Positive

(Colour of HEp-2 cells are in bright green and can be observed clearly)



Intermediate

(Colour of HEp-2 cells is slightly dimmer, but still observable)



Negative

(Colour of HEp-2 cells is very dim and not clear)

(b)

Figure 2.2: Summary of the two main steps in manual IIF test where (a) Step 1: Observation of Cells and (b) Step 2: Classification

There are six common patterns for HEp-2 cell and each of them corresponds to different types of autoimmune disease [2]. The six common patterns are Centromere,

Nucleolar, Homogeneous, Cytoplasmic, Fine Speckled and Coarse Speckled [6, 7] as shown in Table 2.1. The detail of the characteristics of each pattern is shown in Table 2.2. These characteristics are important because they will be used as the main reference for the physicians to diagnosis the HEp-2 cells during IIF test [6].

Table 2.1: Six common staining patterns of HEp-2 cells

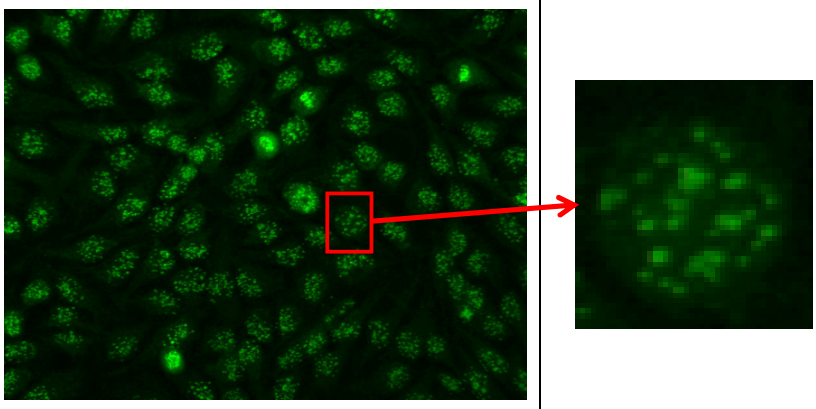
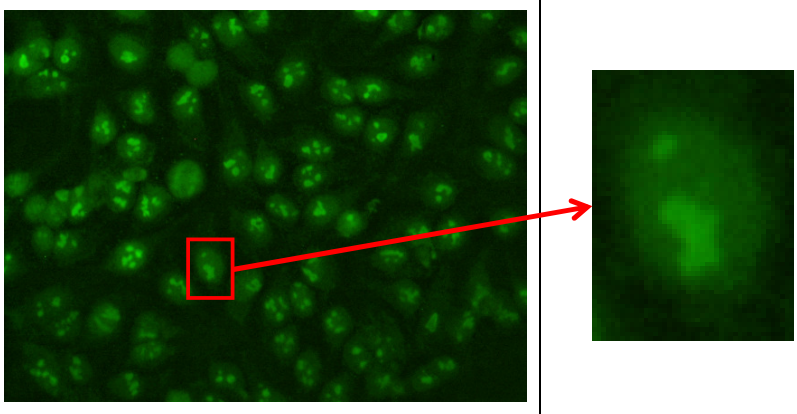
Name of Pattern	**Staining Pattern of HEp-2 Cells (Full)**	**Staining Pattern of HEp-2 Cells (Single Cell)**
Centromere		
Nucleolar		

Table 2.1: Continued.

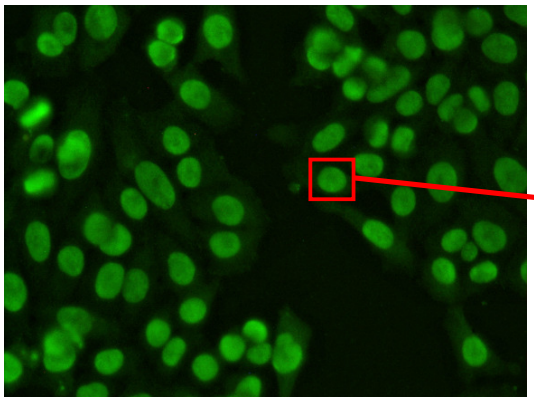
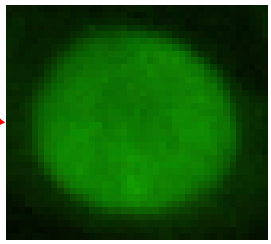
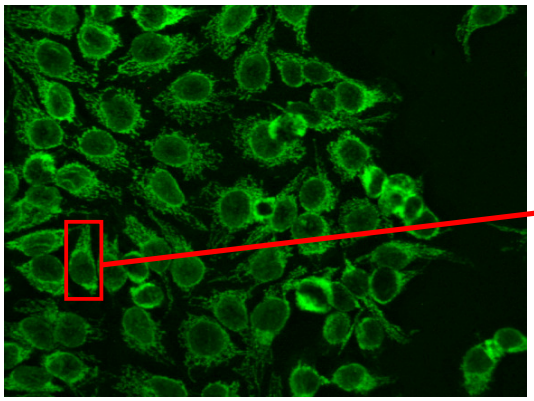
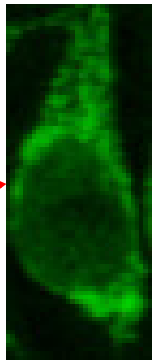
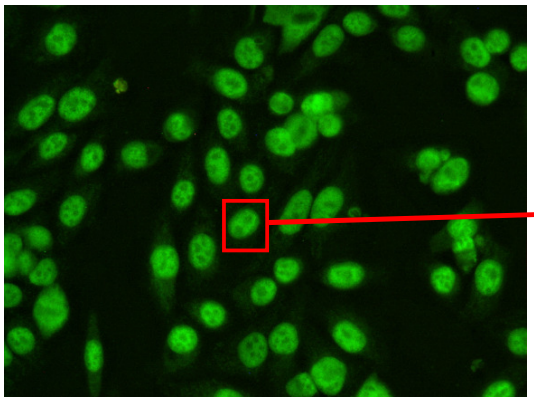
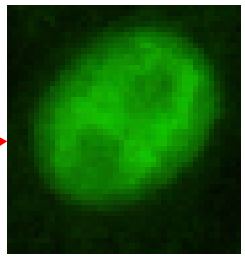
Name of Pattern	Staining Pattern of HEp-2 Cells (Full)	Staining Pattern of HEp-2 Cells (Single Cell)
Homogeneous		
Cytoplasmic		
Fine Speckled		

Table 2.1: Continued.

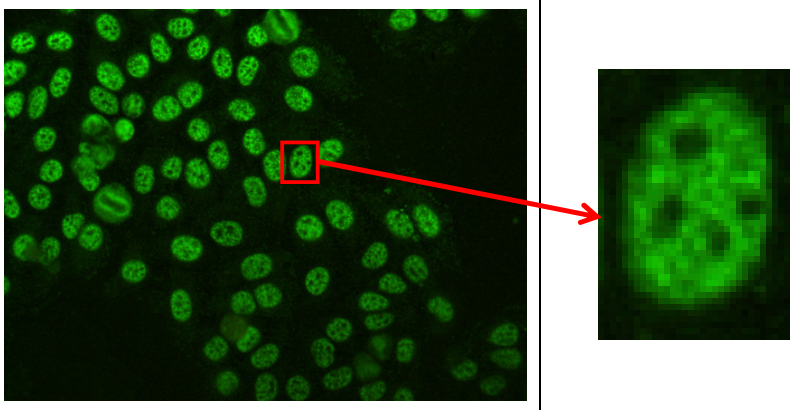
Name of Pattern	Staining Pattern of HEp-2 Cells (Full)	Staining Pattern of HEp-2 Cells (Single Cell)
Coarse Speckled		

Table 2.2: Characteristics of each class of staining patterns of HEp-2 cells

Name of Pattern	Characteristics
Centromere	• Approximately 40-60 discrete speckles are distributed throughout the interphase nuclei [3, 4, 6]. • The speckles are found in the condensed nuclear chromatin during mitosis as a metaphase bar of closely associated speckles [3, 4, 6].
Nucleolar	• Large granules are clustered in the nucleoli of interphase cells which tend towards homogeneity with less than six granules per cell [3, 6]. • Some nucleolar patterns have diffuse cytoplasmic staining in mitotic cells and negative chromosomal region [4]. • Associated with speckled or homogeneous staining of nucleoli and weak speckled or homogeneous staining of nucleoplasm [4].
Homogeneous	• The cells are with diffuse staining of the interphase nuclei and staining of the chromatin of mitotic cells [3, 6]. • The resting cells nuclei have a uniform and diffuse fluorescence appearance [4]. • The condensed chromatin of the mitotic cells has a solid and uniform fluorescence appearance and is more pronounced than the resting cell nuclei [4].

Table 2.2: Continued.

Name of Pattern	Characteristics
Cytoplasmic	• There will be speckles and fine fluorescent fibres running over the length of the cell [3, 6].
Fine Speckled	• The mitotic cells show no staining of the condensed chromosomal regions [4]. • The cells are having a fine granular nuclear staining of the interphase cell nuclei [3, 6]. • The nucleoli may be positive or negative [3].
Coarse Speckled	• The mitotic cells show no staining of the condensed chromosomal regions [4]. • The cells are having a coarse granular nuclear staining of the interphase cell nuclei [3, 6]. • Similar to fine speckled but different in the size of the nuclear [6]. • The nucleoli may be positive or negative [3].

IIF test is always known as the gold standard for detection of autoimmune disease. However, up to date, IIF test has not reached a significant level of automation yet. Some of the researchers claim that IIF methodological procedure has a low level of standardization [2, 4]. For example, some physicians may accidentally use an unsuitable microscope for magnification of the HEp-2 cell images while reading slides. Besides, there is non-conventional method in cutoff dilution of serum and preparation of the slides. IIF test also has the shortcoming in the test results where the test results are mainly depending on the qualification, medical knowledge and experiences of the physicians. This will cause the inconsistency and high variability of the test result [2, 3, 8].

On the other hand, observing the HEp-2 cell images manually through the fluorescent microscope may cause photo bleaching effect where the cells will almost or completely bleached over a short period of time [2, 3]. This will affect the test result of

the HEp-2 cells. The traditional IIF test is also very time-consuming because physicians need to manually inspect the slides with the help of a fluorescent microscope and manually interpret the patterns slides by slides [3]. Because of this, the diagnosis of the ANA will need a long period of time to produce a result which may bring side effect to the patients.

In short, because of these limitations, in the recent decades, researchers have started to move from traditional IIF method to an automated computer aided diagnosis (CAD) systems using digital image processing method. There is a strong need for automated IIF test because of the cost and time saving issue, and also to improve the repeatability of the test result [6]. Further detail on the digital image processing will be discussed in the next section.

2.3 Digital Image Processing

An image can be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial coordinates that represented by horizontal axis and vertical axis respectively. The amplitude of function f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that particular point. The term gray level is always refer to the intensity of a monochrome image. A digital image is defined when all the value of x, y and the amplitude values of f are finite. Converting an image to digital form need to gone through two processes, namely sampling and quantization. Sampling is the process of digitizing the coordinate values of image while quantization is the process of digitizing the amplitude values of image. The result of sampling and quantization form a matrix of real number, this matrix is known as digital image [9].

A digital image will contain M rows and N columns. The size of a digital image will be determined by $M \times N$. Generally, a digital image is made up of a finite number of elements, where each of the elements has a particular location and value. These elements are referred as image elements, picture elements, pels and pixels. Out of them, pixel is the most common term to denote the element of a digital image [9]. The first coordinate value in a digital image will define to be at $(x, y) = (0, 0)$ and the next coordinate value will be at $(x, y) = (0, 1)$ and so on. Pixel of a digital image refers to the coordinate value at the particular location (x, y) of the image. Each pixel will store a value proportional to the light intensity at that particular location. For example, in an 8-bit gray scale image, the pixel will have a value at the range of 0 to 255 [9].

Digital image processing is a process of manipulating and analysis of pictures or images done by a digital computer. Digital image processing has become one of the popular research fields since last century when bigger machines are developed and able to handle high amount of data [10]. The application and usage of digital image processing is very wide, not only limited to medical and biological applications, but also for machine and robotic vision applications [11], geotechnical engineering [12] and astronomy engineering [13]. In this research project, the main focus will be on the medical imaging application. In modern medical field, one of the most effective tools that help physicians to diagnose and analyse sickness and ailment is the medical image processing. With the advancement of the computerized imaging systems for medical treatment, biomedical images are widely used in patient health diagnosis, observation of illness progression, planning treatment and guiding treatment [14].

The formation of medical images come from a few sources, such as X-ray imaging, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound,

and through fluorescence microscope [14]. Each of these methods captures images in different way and the images captured are used for different purpose. For example, X-ray will ionize radiation by sending X-beam shafts through the patient's body to create picture of patient's inside structure. X-ray is always used to investigate the break or fracture of bones in human body. CT is another technique that joins various X-ray projections taken from distinctive edges to create itemized cross-sectional images inside the patient's body. CT is always used for diagnosis of cancers such as lung and liver cancer [14]. MRI is useful in diagnosis of the infected soft tissues inside patient's body. Radio waves are used with an attractive field to make point by point images of organs and tissues in the human body [14].

The next source of medical imaging is ultrasound. Ultrasound utilizes high recurrence sound waves by the ultrasound machine to create images inside human body. The ultrasound machine will send sound waves into patient's body and change the returning sound echoes into an image. Ultrasound imaging is widely used in diagnosing a wide range of conditions affecting the organs and soft tissues of the patient's body, which includes heart and blood vessels [14]. Lastly, fluorescence microscope is another way of getting the medical images for analysis. Unlike others, there is no need to send in any sound waves, radio waves or X-ray into human body. Physicist will take blood samples from the patients, mix the samples with some chemical tester and then capture the image under fluorescence microscope. HEp-2 cell images are the example of the medical image that captured from fluorescence microscope [2].

In digital image processing, for whatever kind of images to be processed and the purpose of the process, the steps to process the images are similar. The first step of digital image processing is image acquisition. For analysing of HEp-2 cell, the images

will be captured using a fluorescence microscope paired with a 50 W mercury vapour lamp and a digital camera. There will be a charge-coupled device (CCD) with squared pixel of equal side to 6.45 μm in the camera. The resolution and colour depth of the images to be captured are depends on requirement, and normally will be stored as bitmap type of images [2]. After the images are captured, the next step will be pre-processing of the image. Image pre-processing is very useful in reducing the unwanted noises in the images as well as improving the image features that are needed for further analysis. Image pre-processing is an optional step in medical image processing because it may decrease the information of the images [2].

After the images are pre-processed, the images will commonly be segmented to detect the region of interest (ROI). Segmentation is an important process to isolate the images into regions with similar properties, for example, texture, shape and image intensity. The segmented images will then proceed to the step of feature extraction. During the process of feature extraction, the features such as texture will be extracted from the images using different method [2]. The details of the method of segmentation and feature extraction in image processing will be discussed in the next two sections. After the features are extracted, normally the features will be stored in database and will be used as the reference library for the classification of the images. Classification of images is the last step in digital image processing. The summarized flow chart of the steps in digital image processing is shown in Figure 2.3.

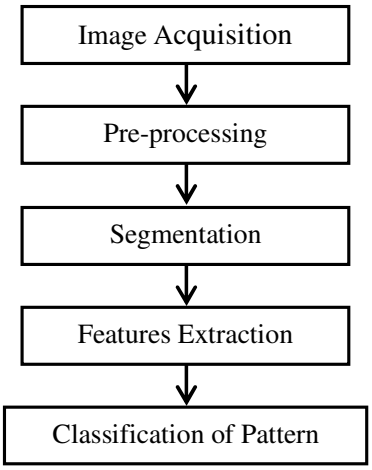


Figure 2.3: Block diagram of the steps in digital image processing

2.4 Segmentation of HEp-2 Cell

According to [7], the appropriate detection of ROI is important before a good classification of HEp-2 cells can be carried out. This is because if the HEp-2 cells are under-segmented, the textural information may be lost. Also, if the background region is incorporated with the HEp-2 cells, extra information may be added to the segmented region of interest, which will misguide the later procedure of the test of ANA [7]. In this research, the ROI is the nuclei of the HEp-2 cells and segmentation process is carried out to locate the HEp-2 nuclei [3]. In this section, the basic concept of segmentation in image processing will be studied and different techniques that are already used by other research works will be reviewed and summarized.

2.4.1 Basic Concept of Segmentation in Image Processing

Segmentation is a process of separating images into coherent parts on the basis of some criteria [14]. It is one of the important steps in digital image processing. Unlike other image processing process where input and output of the process are both images, in segmentation process, the output will not limited to only images but also in the form of important features or attributes extracted from the input images [9]. During the process, the input images will be subdivided into its constituent objects or areas. With different level of problems or requirement, the level for the input images to be subdivided will be different. As long as the objects of interest in the application have been isolated, the segmentation process should stop. This is to prevent extra regions that are unrelated is being segmented. If the images are over-segmented, the final result of the computerized analysis will be affected and the accuracy will reduce [9].

Segmentation in image processing is always targeting and focusing on two types of images, which are nontrivial images and monochrome images. Nontrivial images are normal digital images that have more than one colour. Because of the pixels of the nontrivial images are higher and more complicated, segmentation of nontrivial images is one of the hardest task in image processing [9]. For monochrome images, the segmentation process will be easier because the images are only make up of one colour, for example black-and-white or grayscale. Segmentation in monochrome images are based on two basic properties of image intensity, which are discontinuity and similarity. In taking care of discontinuity, the segmentation process will partition the image based on abrupt changes in intensity, such as edges in that image. For similarity, the segmentation process will partition the image into regions that are similar according to a set of predefine criteria [9].

As mentioned, segmentation in monochrome images is much easier than nontrivial images. Because of this reason, in most of the segmentation research, researchers will first pre-process the images into greyscale images rather than colour images. According to [3], although the researchers agreed that segmentation steps are important and useful in further classification stages, the problems and challenges in segmentation of images are still yet to solve because of the variability in cell appearance due to fluorescence intensities, irregular illumination and staining patterns on the HEp-2 cell images [3]. Because of the challenges, researchers in taking care of HEp-2 cell images already came out with a lot of segmentation techniques to handle different kind of HEp-2 cell images and requirement. In the following section, different type of segmentation techniques will be studied and discussed.

2.4.2 Segmentation Techniques in HEp-2 Cell Image

There are a lot of techniques that can be used to segment the ROI of HEp-2 cell images before extracting feature from the images. In this section, different techniques that are suitable in segmenting of HEp-2 cell images are reviewed and discussed. The techniques that have been used by other researchers are thresholding, watershed, K-means clustering, and Fuzzy C-Means clustering.

2.4.2.1 Thresholding

Thresholding is the most popular and common technique in segmentation of medical images because of its intuitive properties and simplicity in implementing [9]. The concept of thresholding is defined as

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq \mathrm{T} \\ 0 & \text{if } f(x, y) < \mathrm{T} \end{cases} \tag{2.1}$$

where $f(x, y)$ is any input image that composed of light objects and dark background

T is the threshold value

$g(x, y)$ is the thresholded image

In the concept of thresholding, assuming that the objects and the background of the input image can be distinguished into two main groups, the first step is to define the threshold value before separating the image into two groups. After that, define any point in (x, y) of the input image that is bigger or equal than threshold value (T) as '1' or object point, and those points that are less than T as '0' or background point.

There are two types of thresholding segmentation technique, namely global thresholding and local thresholding. For global thresholding, T is constant and the concept of it is similar to the concept mentioned in Equation 2.1. Local thresholding technique is used when global thresholding fails because of the uneven background of the image. Local thresholding will apply a pre-processing step where the threshold is computed to each pixel of the image within a window of radius *r* around it. This is to compensate for the illumination problems and this pre-processing step will be done before applying the global threholding to the image.