Pattern Classification of Human Epithelial Images

by
Mohd Fazlie Bin Mohd Isa
16373

Dissertation submitted in partial fulfilment of
the requirements for the
Bachelor of Engineering (Hons)
(Electrical and Electronic Engineering)

JANUARY 2016

Universiti Teknologi PETRONAS,
32610, Bandar Seri Iskandar,
Perak Darul Ridzuan.

CERTIFICATION OF APPROVAL

Pattern Classification of Human Epithelial Images

by

Mohd Fazlie bin Mohd Isa
16373

A project dissertation submitted to the
Electrical and Electronic Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirements for the
BACHELOR OF ENGINEERING (Hons)
(ELECTRICAL AND ELECTRONIC ENGINEERING)

Approved by,

_____
(Dr. Ibrahima Faye)

UNIVERSITI TEKNOLOGI PETRONAS
BANDAR SERI ISKANDAR
January 2016

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgement, and that original work contained herein have not been undertaken or done by unspecified sources or persons.

_____

MOHD FAZLIE BIN MOHD ISA

ABSTRACT

This project shows an important role to diagnosis autoimmune disorder which is by a comparative analysis on the most appropriate clustering technique for the segmentation and also to develop algorithm for positivity classification. In this project, there are four stages will be used to analyze pattern classification in human epithelial (HEp-2) images. First of all, image enhancement will take part in order to boost efficiency of algorithm by implementing some of the adjustment and filtering technique to increase the visibility of image. After that, the second stage will be the image segmentation by using most appropriate clustering technique. There will be a comparative analysis on clustering techniques for segmentation which are adaptive fuzzy c-mean and adaptive fuzzy moving k-mean. Then, for features extraction, by calculating the mean of each of the properties such as area, perimeter, major axis length, and minor axis length for each images. After that, will implementing a grouping based on properties dataset that has been calculated. Last but not least, from the mean of properties, it will classify into the pattern after ranging the value of mean properties of each of the pattern itself that has been done in classification stage.

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to the following persons who have support and help me to complete my Final Year Project successfully. First and foremost, I would like to extent my heartfelt gratitude to my parents, Mohd Isa Bin Mohd Yusoff and Junianah Binti Hasan for always supporting and give positive advices for me towards completing this project. Moreover, I also would like to extent my gratitude to my supervisors, Dr Ibrahima Faye and Ms. Zazilah Bt May for giving me the most thorough guidance and constant supervision in order to complete this project.

Special thanks to the FYP coordinator, Dr Azrina Bt. Abd Aziz and Dr Norashikin Bt Yahya as they always guide the electrical and electronic engineering's student by providing us with information and reminders. Not to forget to all of the examiners who help me a lot to improve my project through my mistakes. The knowledge they shared to us had opened our mind to make the project more successful. Last but not least, I am highly in debt to the lab technicians, lab supervisors, friends and those who had helped directly or indirectly throughout the period of finishing the project.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background Study

Immune system in our body produce antibodies to help the body against infectious organism. Sometimes, the antibodies misinterprets the body cell as an infectious organism, these antibodies are called autoantibodies. American College of Rheumatology defined that the antibodies which point “normal” protein within the nucleus of a cell is termed antinuclear antibodies (ANA) [1]. Autoimmune disorder of a patient can be examined by using the ANA test as an initial test. [2]. For the time being, indirect immunofluorescence (IIF) is the suggested solution that has been used to detect an ANA. Nowadays, automatic cell detection of images obtained from antinuclear antibodies (ANA) has become highly demand due to the subjective result, time consuming, and bad quality and not standardize result. Furthermore, lack of information and adequately trained physician are not always present also become a crucial issue nowadays.[3] These two realities that enhanced research into the development of automated classification for example a computer-aided diagnosis (CAD) which are able to help the physician regarding the decision.[3]

Regarding the cell detection and classification, there are many method that highly advice to use it to detect and classify the ANA. Classification of ANA in this paperwork is based on the pattern classification in Human Epithelial type 2 (HEp-2) images. The classification itself is about distinguish the images into the pattern such as homogenous, centromere, nucleolar, fine speckled and coarse speckled after implementing the MATLAB algorithm.

1.2 Problem Statement

Referring to previous method that have been done in classifying the pattern of antinuclear antibody in HEp-2, there is a need to have an expertise of physician to support a decision. Due to the limitation from other research, there are a lot of major problems already identified for instance the lack of the quantitative data given to physicians, variety of analysing systems and optics are the problem that the result become variability and subjective. Current implementation requires at least two experts for positivity interpretation will lead to a contradicting opinion between the experts. When physician needs to classify between a low high intensity from an intermediate and between an intermediate from a low intensity they often have a disagreement. Another thing, if the IIF test done by manually it will produce a subjective result and lead to bad quality and not standardize results. The time consuming will also take into account if manually classify.

1.3 Objective

The main objective that need to be accomplished throughout the research process are:

i. To develop algorithm for pattern classification.

The sub-objectives of this paper are:

i. Implement some of adjustment and filtering technique for image enhancement.
ii. Study and do comparative analysis on the most appropriate clustering technique for segmentation of images.
iii. Obtain the percentage accuracy and automated classify the HEp-2 cell images.

1.4 Scope of Study

The project is mainly about the pattern classification of the Hep-2 cell images. It is done by applying MATLAB algorithm on the images in order to distinguishing the cell images into what pattern of it. Commercial image and database will be get it from the hospital or can be taken from the internet which is Mivia Hep-2 Images Dataset. After that, some of adjustment and filtering technique has been applied to enhance the images. Then, there will be a comparative analysis on the most appropriate clustering technique which is for segmentation of images. After that, pattern classification will take part do the programming which is extract the features from the images and then by adding up some algorithm to classify the images by using MATLAB and image processing toolbox. There is a need to develop a supervised approach which does not suffer from the uncertainty.

1.5 Relevancy of Project

This project is significant and relevant to enhance the development of automated classification for example a computer-aided diagnosis (CAD) which are able to assist the physician to make the decision regarding classification of cell pattern. It is also demonstrated the important role to diagnosis autoimmune disorder in which a comparative analysis will be done on the most appropriate clustering technique for the segmentation and also to develop algorithm for pattern classification. Besides that, this project also focus on the implementation of color space model for image enhancement as well as study and do comparative analysis on the utmost suitable clustering technique for segmentation of images.

Furthermore, this research project is relevant for the development of automated classification by determining various features in images which contribute to sample and classifying the HEp-2 cell images to the cell pattern. The information is vital to enable the development of automated classification of HEp-2 images.

1.6 Feasibility of Project

Certain amount of time were allocated to the student to fulfil the project's objectives. This project consists of two stages, the first one is mostly emphasis in literature review, figure out the basic concept and principle of the project. The second stage of the project is focused on making systematic and suitable procedure to conduct the analysis and experiment more effectively. This will ensure the experiment yield less error and give best performance on accuracy of data reading. In addition, image enhancement method which is by using RGB color space model to obtain clearer images and increase efficiency of algorithm, is applied in the experiment. After that, image segmentation method are used to identify the differences between two clustering technique so that the technique which yield better result is identified. The analysis will be continued on the second part of the stages which are features extraction and classification of HEp-2 images by using intensity order pooling and bag of words. The project is feasible within the stipulated time given with appropriate planning. This is to make sure student is able to achieve the objectives within the short time frame given.

CHAPTER 2

LITERATURE REVIEW AND THEORY

Autoimmune disorder in HEp-2 cells is the main part of diagnosis. However, performances of current systems available in literature are not satisfying. Based on the previous research, the method depend on specialists to observe HEp-2 slides via the fluorescence microscope, which suffers from a number of shortcomings like being subjective and labor intensive.[4]

The most vital step to increase efficiency of algorithm for segmentation by picking a suitable color space model, it also produce a significant value. Ketenci et al. presented a comparative analysis on performance for skin segmentation by using simple color space of 2D Gaussian color model. There are several types of color space that identify as a universal, there are RGB, HSV, YIQ, and YUV. RGB color information consists of red (R), green (G) and blue (B), while hue (H) and (S) saturation is for HSV. For the other two which have in-phase (I) and quadrature (Q) for YIQ, while for U and V respectively consists of blue and red chroma for YUV. For respective pixel, probability value was calculated with the component of color information. The value obtained in the range of (0, 1) represent the degree of skin color. Lastly, the detection of skin and non-skin part, uses the chosen threshold values by Otsu method. ROC curves are used in analysing the performance of 2D Gaussion Color Model in RGB, HSV, YIQ and YUV. Hence, the algorithm provides an excellent result in color space. As a result, RGB color system give the best result performance compared to others such as YUV, YIQ, and HSV. There is a similarities of curves between YUV and HSV, however YUV consists large area under curve compared to the HSV. [5]

One of the difficult task for image analysis is image segmentation. For numerous utilization, a lot of segmentation algorithm have been develop. For most current segmentation algorithm, dissatisfying results are experienced in some cases.

Isa et al. propose a new composition of traditional moving k-means for segmentation of image which are termed adaptive fuzzy moving k-mean, adaptive moving k-mean and also fuzzy moving k-mean. After conducting the analysis, it shows that the new composition algorithm which is adaptive fuzzy moving k-mean are not affected much by the first step of grouping value and also noise. Other than that, performance of this new composition algorithm is better than other clustering algorithm and it also can surpass the problem occur in algorithm that was discovered before [6]. Nowadays, for the segmentation of image, the clustering method "Fuzzy C-Means" (FCM) is commonly used. When segmenting images with noise, FCM produces a good result. However, for images which have been tainted by noise or consists of imprecise edges, the segmentation cannot be done exactly. Because of that, Ayech et al. introduce alternative technique that want to overcome this trouble by using adaptive distance of FCM. As a result, the alternative technique of FCM shows that there is a significant enhancement for performance compared to traditional FCM [7]. The downside of traditional FCM is that repetitive test need to be conducted for the image segmentation. To overcome this problem, a new modification of FCM which is by using adaptive spatial has been designed by Yu et al. [8]

Hobson [9] presented a benchmarking platform of classification anti-nuclear antibodies Hep-2 image by applying CAD system which is not just simple but also effective. This CAD system was inspired from the recent success research in the object classification domain which is represented by Cell Bank. The information which are utilized by this paper is separated from both interphase and mitotic cells. The classifier which consists of individual HEp-2 cell image extraction, followed by classification of cell cycle, then continue with the cell image descriptor extraction and classification and lastly ANA image descriptor extraction and classification. 66.4% and 53.0% mean class accuracy (MCA) for interphase and mitotic respectively. Soda and Iannello [10] introduced a method to classification on Hep-2 images which is a bag of visual words (BOVW) access for staining traits of cytoplasmic and centromere. This approach shows complicated picture contents by not using the step of segmentation. It looks for region of interest conveying the most helpful data, and then practices a BOVW approach such as local information extraction and visual vocabulary construction. Regarding local information extraction, local data from the Hep-2 images were divided by using Well-known Scale-invariant Feature Transform (SIFT) descriptor. Based on

visual vocabulary, *k*-means algorithm has been use to group the set of ROIs separate from practice images. Though over 98% accuracy was achieved to discriminate centromere and cytoplasmic with other patterns, performance of the system dropped to as low as 64.2% when all 6 patterns were considered.

Larsen et al. propose a design for automatic analysis of IIF images of Hep-2 cells into others staining traits character. This approach is to taking second order picture structure at different scales pertaining to a new texture part termed shape index histograms. The donut-shaped pooling regions is used to changing sizes when collecting histogram input of spatial decomposition.[11] The system achieved around 83.53% accuracy on the test set. Among the many staining pattern which can be observed, Figure 2.1 shows the staining six patterns that are relevant to diagnosis purpose. These include centromere, homogeneous, nucleolar, coarse speckled, Golgi, nuclear membrane (NuMem).

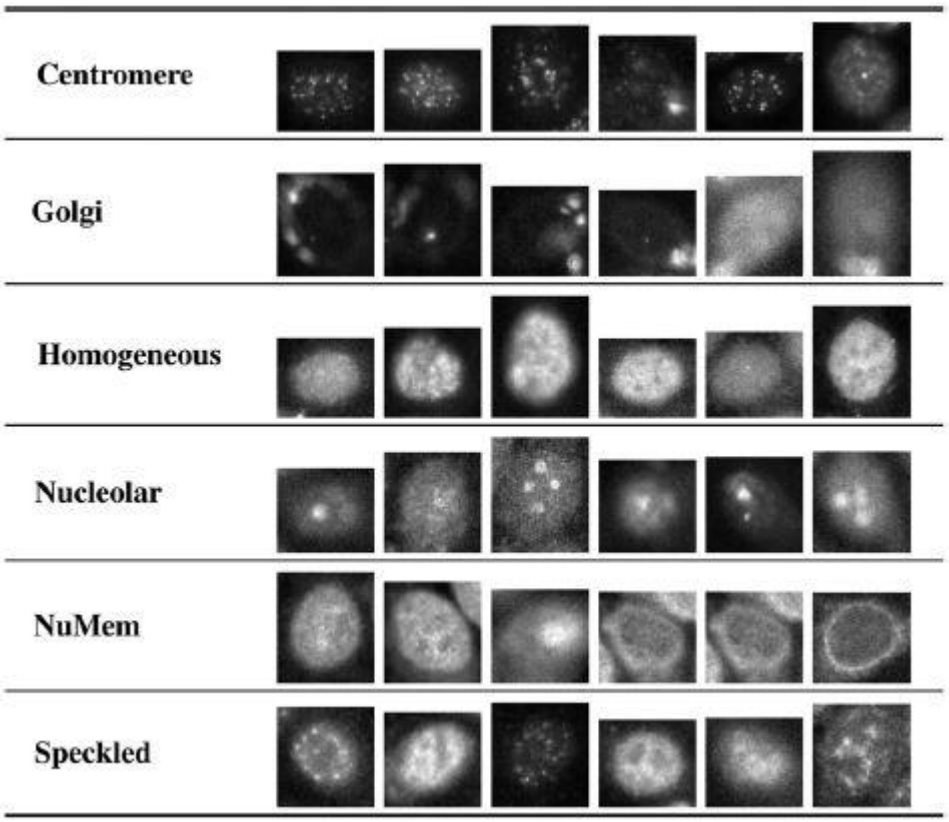

Figure 2.1 Example of the six others staining pattern characters

Propose in [4], by applying intensity order pooling gradient feature and bag of words (BOVW), Hep-2 can be classify. The pooled feature will be invariant without requirement orientation estimation and when the gradient features are combine according to the intensity of orders of local grid points. 100% image level accuracy was achieved on the SZU dataset. Ghosh and Chaudhary [12] perform an experiment by using combination of features like Histogram of Oriented Gradients (HOG) features, Region of Interest (ROI) features, and Speed-Up Robust Features (SURF), texture features. Combination of ROI, texture and HOG using SVM classifier gives the best performance based on the experimental results. 60% accuracy was reported on the test set. The literature also use other features such as eccentricity, shape [13], granulometry [14], and euler number [15]. The top competitor reported in [16] uses Co-occurrence of Adjacent LBP (CoALBP) for feature extraction and SVM for classification, 68.7% accuracy was achieved.

CHAPTER 3

METHODOLOGY AND PROJECT WORK

3.1 Experimental Work

3.1.1 Propose Method

Figure 3.1 Overall system block diagram

Based on the Figure 3.1, there are the overall system block diagram that have to fulfill in order to classify the cell.

1) The cell image database will be taken from the internet datasheet.
2) Next, for the image enhancement there will be a changing from RGB to grayscale. After that, the author apply some of the adjustment and filtering technique to get clearer image.
3) Step will be following by image segmentation part which is the image will be divided from the background. There will be a comparison method of clustering technique which is adaptive fuzzy c-mean and adaptive fuzzy moving k-mean used for image segmentation.
4) By using features extraction step, the information will be obtain from original data and show the information in narrow scope. The characters of cell that have autoantibodies are:

i. Fluorescence staining at cell nucleus.

ii. Give fluorescence intensity information either positive, negative or others.

5) By using proposed method, the cell will be classify into some of pattern based on fluorescence intensity that can be used for diagnosis purpose.

3.1.2 Tools and Software

i. IIF Hep-2 Images
ii. MATLAB and Image Processing Toolbox
iii. Microsoft Office

3.1.3 Design Approach

Based on the literature above, the author decided to apply the existing method which is by implement RGB color space for image enhancement then do the clustering technique for segmentation and lastly applying intensity order pooling pertaining to gradient feature and bag of words for classifying HEp-2 [4] from the previous research with the improvement to overcome their limitation. The approach will be explain in more detail.

a) RGB Color Space Model

As the name implies, RGB model uses 3 color elements that consist of R, G and B. The human eye functionality uses the same concept which these three color schemes are utilized. In the retina, there are three types of cone that correlate to the specific light spectra (RGB) [17]. Hence, the when intensity of red, green and blue increases, more visible colors can be acquired. There are many applications which uses the RGB mixing model such as computer graphics and in cameras.

RGB space can be displayed in a 3D Cartesian coordinate system in cubical shape (Refer Figure 3.2). The white and black color are situated in the corner and the origin respectively while primary and secondary colors are placed at the vertices. For the proposed algorithm, 24-bit RGB is needed and each components have an 8-bit depth. 16,777,216 will be the total number of available colors. A single byte of the component will have values of unsigned integers in the range of (0,255)

Basically, original images are represented in the 24-bit RGB format. Even so, RGB will have to be floating-point integers ranging (0, 1) for classical conversion algorithm, that causes type-casting and operation of division. In the basic 24-bit format, the channels are prepared for the proposed algorithm. As both of the algorithms worked independently, the order of the components are insignificant.

Despite the excellent functionality, RGB has a few drawbacks. Cannot be interpreted with human intuition. Luminance are chrominance properties are clumped together thus, making it hard to classify. Color differences are not able to be spotted properly as Euclidean distance cannot be implemented [18]. Last but not least, the components are mostly related to each other and also responds greatly to illuminance changes and noise. [19]

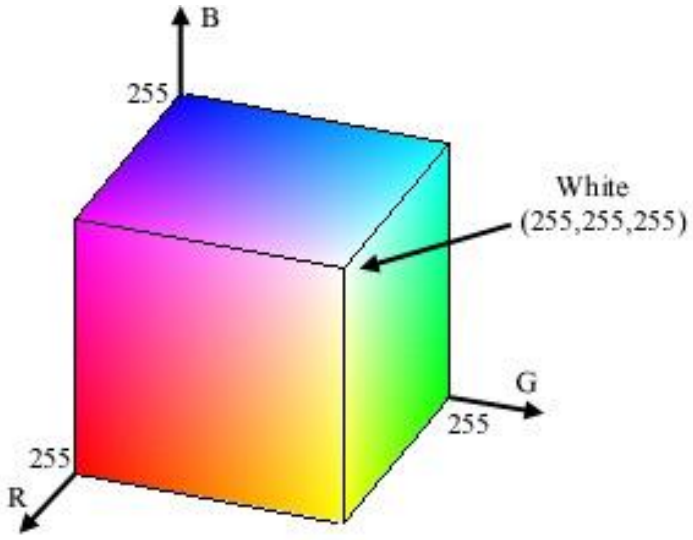

Figure 3.2 RGB Colour Cube

b) Segmentation Method

Segmentation is separate image form background. Also, is a process of splitting an image into number of sub images or extracting the necessary portions from the image. In these paper, the proposed technique is actually do the comparison between two clustering technique which are adaptive fuzzy moving k-mean (AFMKM) and also adaptive fuzzy c-mean (AFCM).

1) Adaptive Fuzzy Moving K-Mean (AFMKM)

$$c_j = \frac{1}{n_j}\sum_{i \in c_j} v_i \tag{1}$$

$$f(c_j) = \sum_{i \in c_j} \left(\left\| v_i - c_j \right\| \right)^2 \tag{2}$$

$$f(C_s) \geq \alpha_a f(C_l) \tag{3}$$

$$c_s = \frac{1}{n_s}\sum_{i \in c_s} v_i \tag{4}$$

$$c_l = \frac{1}{n_l}\sum_{i \in c_l} v_i \tag{5}$$

$$\alpha_a = \alpha_a - \alpha_a / n_c \tag{6}$$

$$f(C_s) \geq \alpha_b f(C_l) \tag{7}$$

$$\alpha_b = \alpha_b - \alpha_b / n_c \tag{8}$$

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}} \tag{9}$$

$$f(C_s) < \alpha_a f(C_l) \quad and \quad m(C_{sk}) > m(C_{lk}) \tag{10}$$

In k-mean clustering, there is an issue need to solve which is *N* data need to clustered into n_c centers. Assume c_j as *j*-th data and v_i as *i*-th middle with fixed first value where *j = 1,2,...,*n_c and *i* = 1,2,…,*N*. Basically, by using Euclidean distance, all the data can be appointed to the nigh center. Referring to (1), it will calculate for each center the new position. The fitness for each cluster can be can be calculated by using (2).

The interrelation among the centers must follow certain condition, after applying the fitness calculation process. The condition is as follow (3), where α_a a small constant with first value is same with α_0 and α_0 a small constant number range 0 - $\frac{1}{3}$. For the cluster that has smallest fitness number is represented as C_s, and the largest fitness number is C_l. In (4) and (5), the equation is to calculated the new position of C_s and C_l due to an unfulfilled condition where the C_l has smaller value compared to C_s. By using eq. (6), the value of α_a can be updated. Hence, to enhance the clustering technique, another requirement has been introduced by duplicating the whole process until (7) is earned where α_b small constant with first value same with α_0. For every iteration α_a is reboot to α_0 and α_b is renew based upon (8).

The purpose of (9) is to allow every data member to be appoint together become more than single class by grade of membership, where typical value of q as the uncertainty exponent is 2, d_{jk} is length from dot *k* to other cluster center *j* and d_{ik} is length from dot *k* to ongoing cluster center *i*. The starting interrelation between the centers (3) and (7) is altered to achieve eq. (10) after done with determining membership for every data and implement the calculation operation of fitness by using eq. (2). In eq. (10), $m(C_{sk})$ and $m(C_{lk})$ are the smallest and largest center based on membership value of dot k. This proposed technique is termed Adaptive Fuzzy Moving K-Mean (AFMKM) that every member will be select to the suitable center after appoint to the fuzzy membership function (9) and fitness function (2) until (10) is accomplished to the nearest center for all the data.

2) Adaptive Fuzzy C-Mean (AFCM)

$$J = \sum_{i=1}^{C}\sum_{j=1}^{n}(u_{ij})^{m} D\ (x_j, v_i) \tag{11}$$

$$u_{ij} = \left(\sum_{k=1}^{C} \left(D(x_j, v_i) / D(x_k, v_i) \right)^{\frac{1}{m-1}} \right)^{-1} \tag{12}$$

$$v_i = \sum_{j=1}^{n}(u_{ij})^{m}\ x_j \Big/ \sum_{j=1}^{n}(u_{ij})^{m} \tag{13}$$

The purpose of eq. (11) is to minimize the objective function of FCM algorithm, where *C* is the value of pixels will be clustered, *n* the cluster number and *m* the exponential weigh of membership degree.

i. u_{ij} appear as membership degree of *j*-th object in the *i*-th cluster,

ii. v_i appear as the *i*-th cluster center,

iii. *D* appears as length metric which determine the equality among object and a cluster center,

iv. $m \geq 1$ the degree of fuzzyfication.

By measuring the angle of *J* with respect to u_{ij}, the membership degree of x_j to the *i*-th cluster can be identified. Hence, these membership degrees can be obtain in eq. (12). By measuring the angle of *J* with respect to v_i, the cluster center of v_i, *i*:1..C can be identified. Hence, these center can be obtain in eq. (13). The following steps are the outline for algorithm.

i. Fix and compute the cluster value and centers by arbitrary points from data set.

ii. Renew the membership degrees by using eq. (12).

iii. Renew the centers by using eq. (13).

iv. Rerun steps 2 and 3 until convergence.

When there is a change in term of membership values is below than a given threshold, the convergence of this algorithm can be reached.

c) Feature Extraction

Feature extraction is the process obtain relevant information from original data and represent that information in a lower dimensionality space. In the MATLAB algorithm, there are four features that has been extracted from each of the image which are area, perimeter, major axis length, and minor axis length. By calculating the average value and grouping for each features of training dataset, the images can be classify into some of the pattern of cell which are homogenous, centromere, nucleolar, fine and coarse speckled.

Intensity order pooling based features

Accuracy for classification is the important part to determine due to the various features such as orientations and rotationally invariant. For each ROI, techniques like SIFT measure the primary orientation and it also utilize this measurement for orientation normalization. Characteristic in a naturally rotation constant way such as ring structure will be extract by using another technique. By the way, to extract that character is about cost of structural data and character inequity. Consequently, there are many methods that are using the technique of reference orientation for rotation the ROI.

Thus, extracting rotationally same characteristic is important for orientation measurement. Based on [20], the respective points can be matched correctly by using the approximation error to be less than 20°. Unfortunately, estimation algorithm around one third provide approximation error more than 20° for ongoing orientation. For local representation, it is important to implement a rotationally invariant descriptor with no condition of orientation approximation as the orientation variance are very common among HEp-2 cells. The intensity order pooling which is recommended by a lately descriptor is a good choice [4].

d) Classification

For the classification, the cell image will be classified and regroup based on the training dataset properties that already obtain in the feature extraction part.

Bag of words

Suggested in [21], BOW was first proposed for visual categorization and it also was originated from the concept of distinguish a text document by measuring the most recurring words. Categorized any image toward a number of blocks and represent them by local characteristic is the total concept of BOW for image representation. By using learned dictionary, it can be shown by the coding of all block when the test image is divided into a number of small block. Analyzing system of cell image by applying BOW will consists of a training and test method. Training sequences are as follow in Figure 3.3:

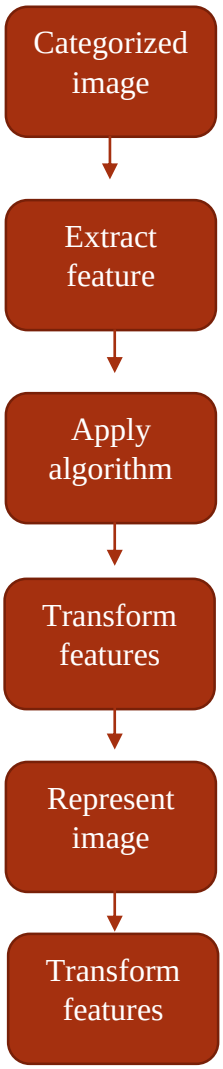

Figure 3.3 The training process

3.2 Project Flow Chart

Below is the process flow for the research project in sequence to complete the objectives. Refer Figure 3.4.

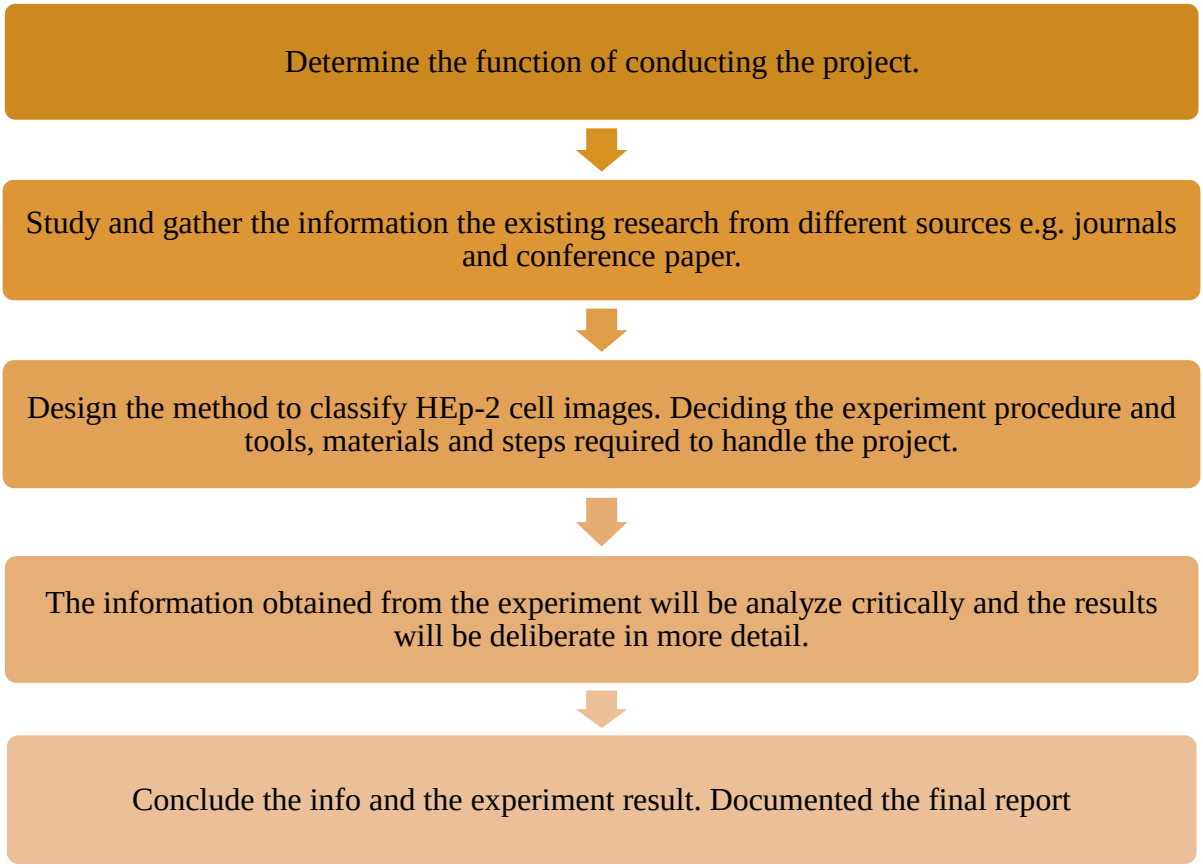

Figure 3.4 Process flow chart for the research project

3.2.1 Project Activities

1. Determine the function of conducting the project.
    i. Do a research on HEp-2 cells (the cause, effect, features and solution to overcome).
    ii. Do a research on digital image processing.

2. Study and gather the information the existing research from different sources e.g. journals and conference paper.
    i. Read, understand and analyze about the research paper on digital image processing
    ii. Find the appropriate and suitable technique available (pros and cons)

3. Design the method to classify HEp-2 cell images. Deciding the experiment procedure and tools, materials and steps required to handle the project.
    i. Preprocessing (Image Enhancement) – Applying RGB color space model.
    ii. Image Segmentation - Performing algorithm from different clustering technique for improved segmentation of HEp-2 cells image.
    iii. Feature Extraction - Performing algorithm from intensity order pooling technique in favor to be able to extract a clearer feature of HEp-2 cells.
    iv. Classification – Performing algorithm from bag of words method for a greater percentage accuracy classification of HEp-2 cells.

4. The information obtained from the experiment will be analyse critically and the results will be deliberate in more detail.
    i. Preprocessing
    ii. Segmentation
    iii. Feature Extraction
    iv. Classification.

5. Conclude the info and the experiment result. Documented the final report.
    i. Make a meeting with supervisor and do a report to update progress of the project.
    ii. Analyses the result obtained and make modification if needed.

3.3 Gantt Chart and Key Milestone

Table 3.1 shows the Gantt chart which is the scheduled project activities and key milestone in order to complete the project.

Table 3.1 Gantt chart and Key Milestones for FYP1

No	Elements	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Choose a topic														
2	Research Study and Literature Review														
3	Preliminary Proposal Report														
4	Final Proposal Report														
5	Preliminary for Proposal Defence														
6	Proposal Defence														
7	Experimental Work														
8	Draft Interim Report														
9	Final Interim Report														

Table 3.2 Gantt chart and Key Milestones for FYP2

No	Elements	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Research on algorithm for image segmentation														
2	Implement and modified feature extraction algorithm														
3	Research on algorithm for feature extraction and classification														
4	Implement and modified classification algorithm														
5	Progress Review Report Submission														
6	Pre-SEDEX														
7	Submission of Draft Final Report														
8	Technical Paper Report Submission														
9	Viva														
10	Dissertation Report Submission														

Process **Propose Milestone**

CHAPTER 4

RESULT AND DISCUSSION

The results show below still undergo testing phase. The procedures for the method for this project is not finalized yet, hence nothing much can be deliberate and explained. In this project, one of dataset has been used for testing purpose which is ICPR HEp-2 Cell Classification Contest Dataset (ICPR). For the first part of this project, image enhancement will take part as initial step in order to classify the cell image. After that, will followed by image segmentation which conduct a comparison between two methods which are adaptive fuzzy c-mean and adaptive fuzzy moving k-mean. Then, for the feature extraction, the author will extract the value data from the image by using shape properties extraction. Lastly, for the classification, the method used is data grouping based on analysis of shape properties. The figure below shows that the step that has been apply in the MATLAB algorithm

1) Homogeneous

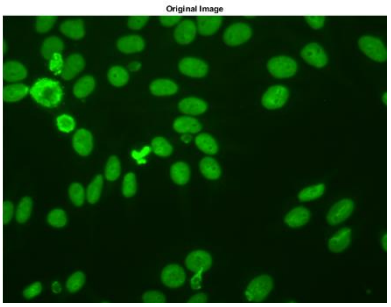

Figure 4.1.1 Original Image

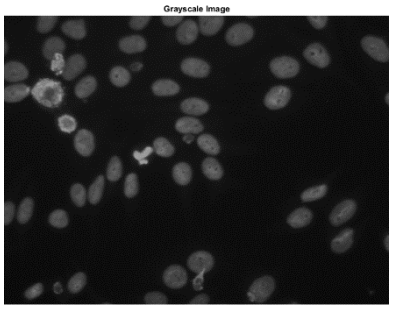

Figure 4.1.2 Grayscale Image

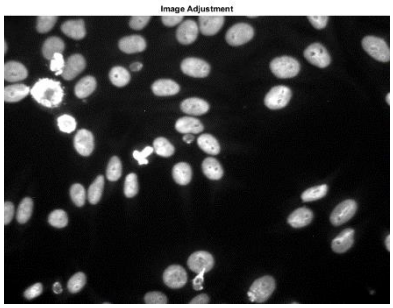

Figure 4.1.3 Image Adjustment

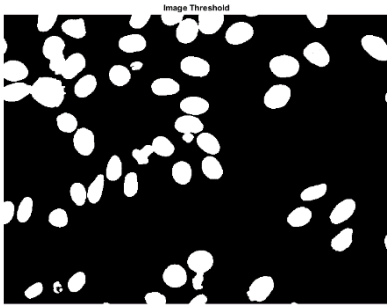

Figure 4.1.4 Image Threshold

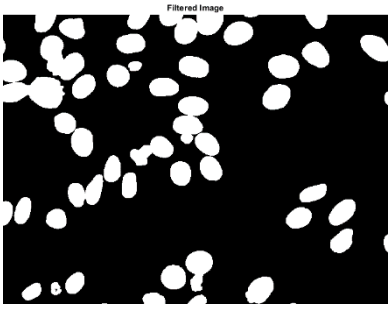

Figure 4.1.5 Filtered Image

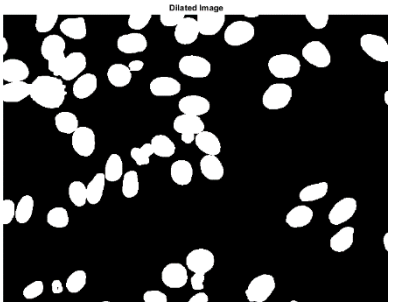

Figure 4.1.6 Dilated Image

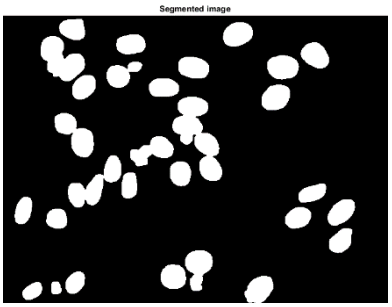

Figure 4.1.7 Segmented Image

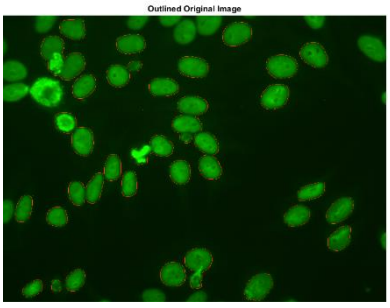

Figure 4.1.8 Outline Original Image

2) Centromere

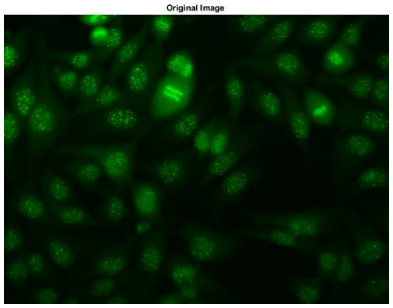

Figure 4.2.1 Original Image

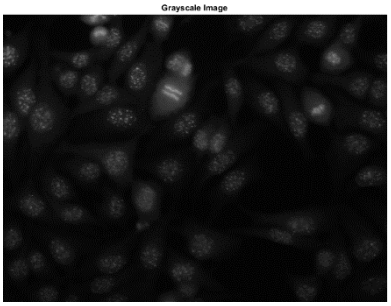

Figure 4.2.2 Grayscale Image

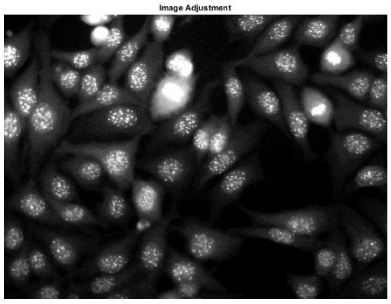

Figure 4.2.3 Image Adjustment

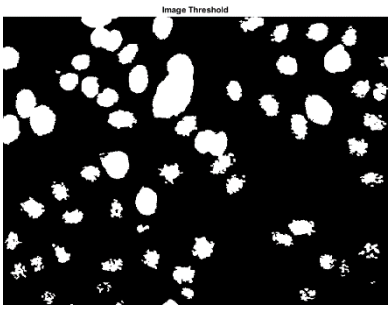

Figure 4.2.4 Image Threshold

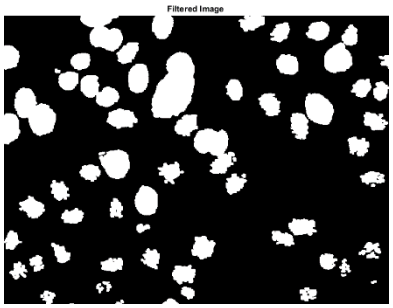

Figure 4.2.5 Filtered Image

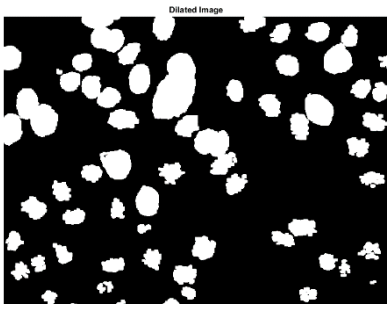

Figure 4.2.6 Dilated Image

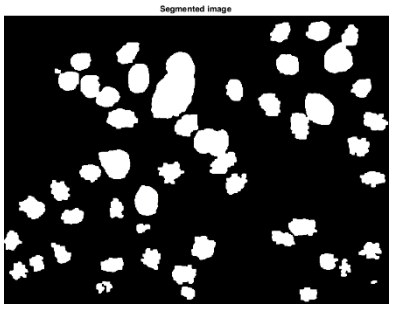

Figure 2.2.7 Segmented Image

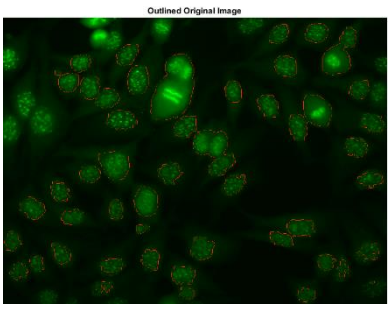

Figure 4.2.8 Outline Original Image

3) Nucleolar

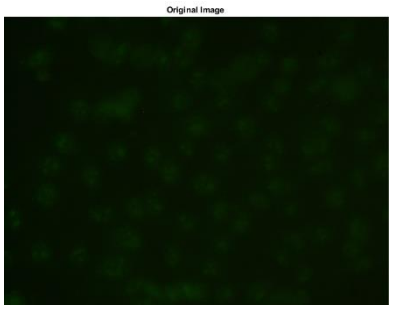

Figure 4.3.1 Original Image

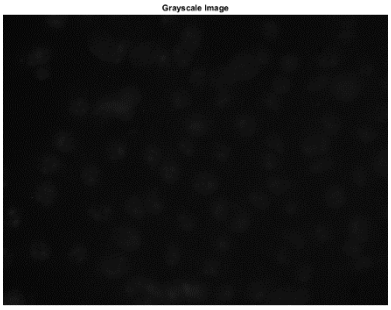

Figure 4.3.2 Grayscale Image

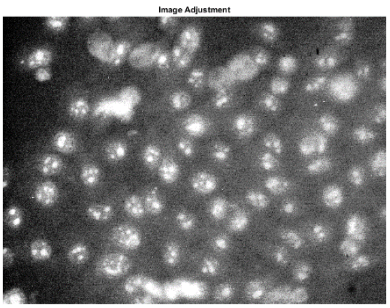

Figure 4.3.3 Image Adjustment

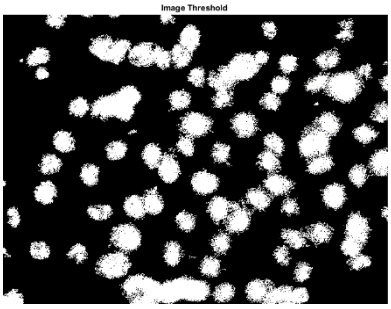

Figure 4.3.4 Image Threshold

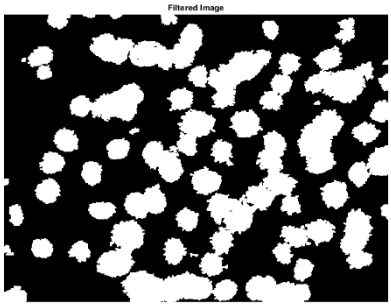

Figure 4.3.5 Filtered Image

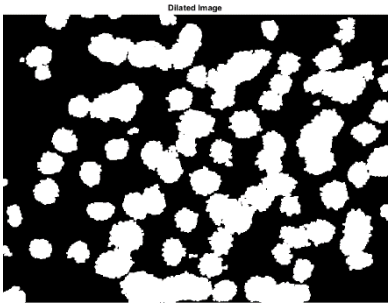

Figure 4.3.6 Dilated Image

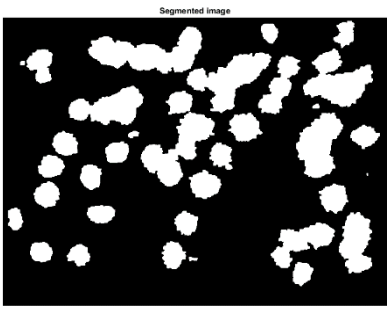

Figure 4.3.7 Segmented Image

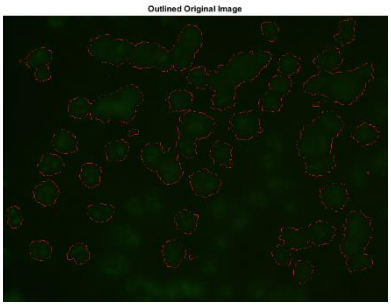

Figure 4.3.8 Outline Original Image

4) Fine Speckled

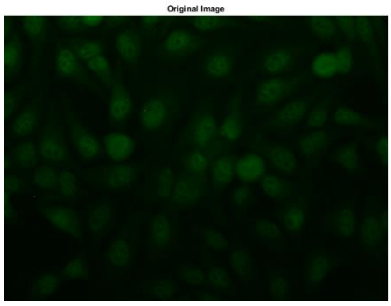

Figure 4.4.1 Original Image

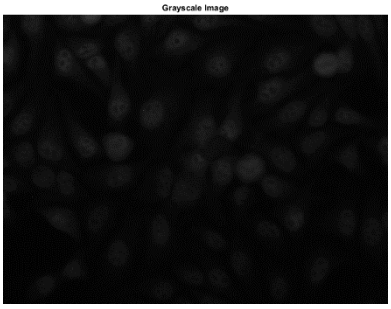

Figure 4.4.2 Grayscale Image

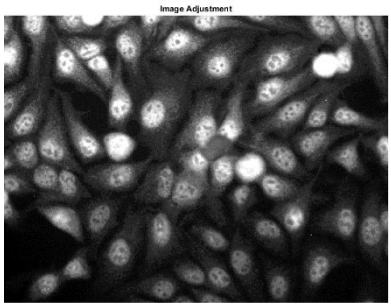

Figure 4.4.3 Image Adjustment

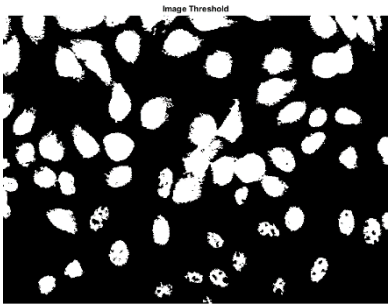

Figure 4.4.4 Image Threshold

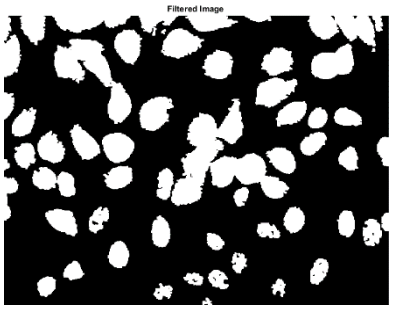

Figure 4.4.5 Filtered Image

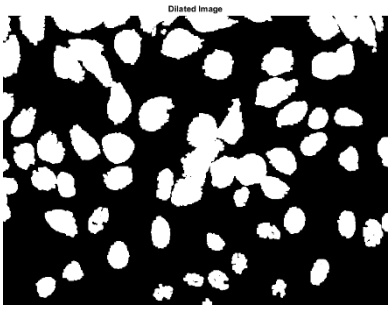

Figure 4.4.6 Dilated Image

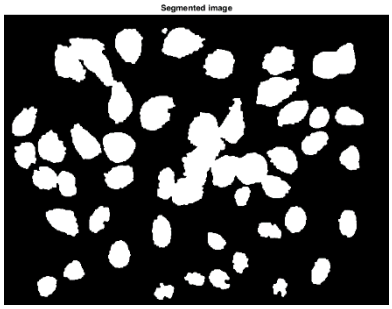

Figure 4.4.7 Segmented Image

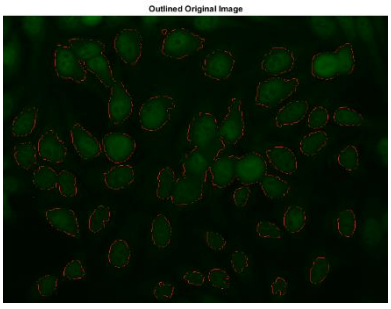

Figure 4.4.8 Outline Original Image

5) Coarse Speckled

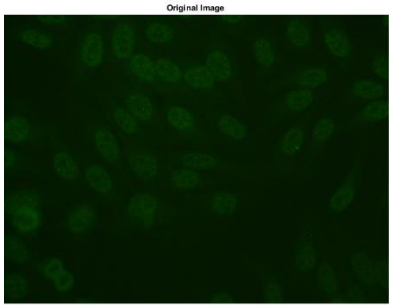

Figure 4.5.1 Original Image

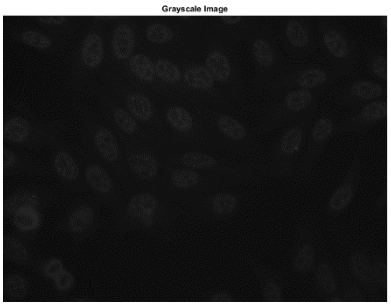

Figure 4.5.2 Grayscale Image

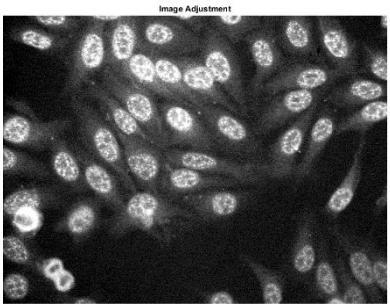

Figure 4.5.3 Image Adjustment

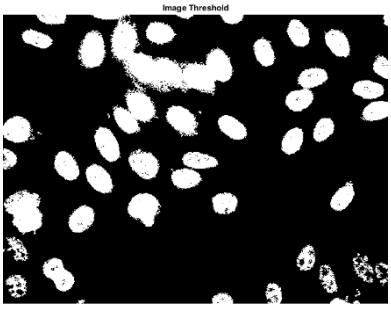

Figure 4.5.4 Image Threshold

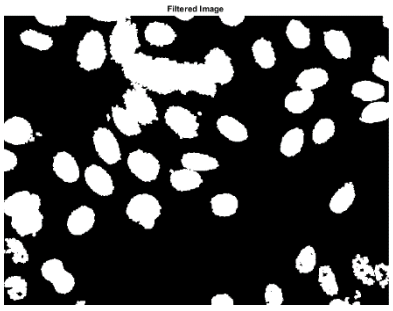

Figure 4.5.5 Filtered Image

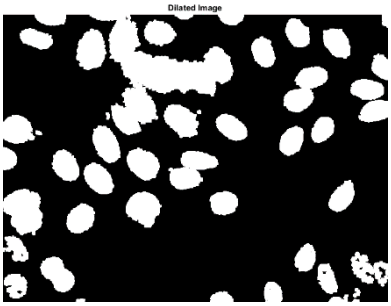

Figure 4.5.6 Dilated Image

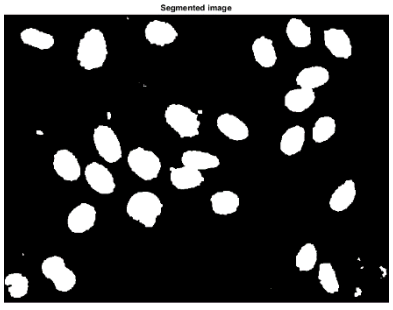

Figure 4.5.7 Segmented Image

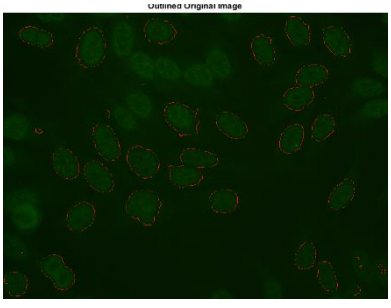

Figure 4.5.8 Outline Original Image

The images displayed above are the step get it from the programming MATLAB. The Figure 4.1.1, 4.2.1, 4.3.1, 4.4.1, 4.5.1 are the original images when apply the algorithm on MATLAB which is read the image from the file. For the enhancement of image, the image will undergo a process which convert it to the grayscale image, and apply some of adjustment and filtered in order to get clearer images. Refer Figure 4.1.2 - 4.1.6, 4.2.2 – 4.2.6, 4.3.2 – 4.3.6, 4.4.2 – 4.4.6, 4.5.2 – 4.5.6. For the image segmentation, refer Figure 4.1.7, 4.2.7, 4.3.7, 4.4.7, 4.5.7. Feature extraction has been conducted a method which takes the mean value for each properties for each image. Assume, if there are 28 images so it will be 28 mean value for each of properties which are area, perimeter, major axis length, and minor axis length. Continue with grouping pattern based on originality dataset, then the author review the value for each properties and pattern and try to get the range value in order to classify the cell.

Table 4.1 Properties in the features extraction part

Pattern	Images number	Properties			
		Area	Perimeter	MajorAxisLength	MinorAxisLength
Homogeneous	1	7.3382	0.3331	0.1119	0.0770
	5	6.4467	0.3362	0.1080	0.0737
	18	6.3535	0.3097	0.1118	0.0669
	21	3.9069	0.2500	0.0840	0.0567
	22	5.8306	0.3264	0.1053	0.0674
Fine Speckled	2	7.4434	0.3900	0.1214	0.0729
	9	6.9649	0.3328	0.1123	0.0759
	15	8.9816	0.4169	0.1242	0.0736
	23	6.0753	0.3061	0.1057	0.0669
Centromere	3	1.5623	0.1781	0.0531	0.0354
	7	3.7279	0.2846	0.0786	0.0525
	13	4.7235	0.2821	0.0885	0.0610
	14	4.6530	0.2869	0.0959	0.0587
	16	2.7796	0.2551	0.0726	0.0475
	19	2.1299	0.2129	0.0627	0.0417
Nucleolar	4	1.0459	0.0501	0.0145	0.0080
	8	1.4315	0.1441	0.0486	0.0330
	20	4.6335	0.2961	0.0910	0.0575
	24	7.3461	0.3796	0.1137	0.0706
Coarse Speckled	6	3.8375	0.2374	0.0796	0.0547
	10	8.0567	0.3762	0.1162	0.0760
	11	5.6165	0.3084	0.0957	0.0526
	12	7.3192	0.3977	0.1194	0.0711
	17	6.2897	0.3038	0.1067	0.0707

Table 4.2 Patten classification based on adaptive fuzzy moving k-mean (AFMKM)

	Ho	Ce	Nu	FS	CS
Ho	97.1%				
Ce		96.8%			
Nu			94.3%		
FS				73.5%	
CS					65.4%

Table 4.3 Patten classification based on adaptive fuzzy c-mean (AFCM)

	Ho	Ce	Nu	FS	CS
Ho	89.4%				
Ce		95.0%			
Nu			91.8%		
FS				85.2%	
CS					79.7%

Ho: Homogenous; Ce: Centromere; Nu: Nucleolar; FS: Fine Speckled; CS: Coarse Speckled.

Table 4.1 shows the properties that has been extracted from the images for feature extraction part. There are four type of properties which are area, perimeter, major axis length and minor axis length. These properties has been used in order to classify the cell into pattern for pattern classify part. Table 4.2 and 4.3 are the percentage accuracy that have been go through in MATLAB algorithm, for two comparison clustering method which are Adaptive Fuzzy Moving K-Mean (AFMKM) and Adaptive Fuzzy C-Mean (AFCM). In the table shows that AFMKM clustering technique have better accuracy than AFCM for 3 patterns which are homogenous, centromere and nucleolar. While, for other two which are coarse and fine speckled, AFCM give better accuracy compare to AFMKM.

CHAPTER 5
CONCLUSION AND RECOMMENDATION

Conclusion, this project shows an important role to diagnosis autoimmune disorder which is by a comparative analysis on the most appropriate clustering technique for the classification of positivity and also to develop algorithm for positivity classification. A disease related to the connective tissue disease on human. This project would enhance the analysis of the disease and hopefully will lead to a solution for a cure.

The author hopes it can bring more benefits in the development of automated classifying ANA. Moreover, the capability of final year student with the support from supervisor and coordinator and the time consuming need to be considered in this research project. It is hoped for the achievement of the research when gaining of tool and materials required for analysis runs smoothly.

The source of data is from internet, Mivia HEp-2 Image Datasheet. The proposed approach was fully tested using publicly available which is International Conference on Pattern Recognition (ICPR) Hep-2 Cell Classification Contest Dataset. The ICPR dataset contains 28 images which consist of Centromere, Homogeneous, Nucleolar, Coarse speckled and Fine Speckled. Further research can be done by applying the best feature extraction and classification technique in order to achieve high value of accuracy such as Intensity Order Pooling and Bags of Words.

REFERENCES

[1] American College of Rheumatology. (2015, 21 October). *Antinuclear Antibodies (ANA).* Available: http://liv.ac.uk.libanswers.com/faq/49511

[2] A. A. C. Chemistry. (2014). *ANA, Antinuclear Antibody* Available: https://labtestsonline.org/understanding/analytes/ana/tab/test/

[3] P. Soda and G. Iannello, "Aggregation of Classifiers for Staining Pattern Recognition in Antinuclear Autoantibodies Analysis," *Information Technology in Biomedicine, IEEE Transactions on,* vol. 13, pp. 322-329, 2009.

[4] L. Shen, J. Lin, S. Wu, and S. Yu, "HEp-2 image classification using intensity order pooling based features and bag of words," *Pattern Recognition,* vol. 47, pp. 2419-2427, 2014.

[5] S. Ketenci and B. Gencturk, "Performance analysis in common color spaces of 2D Gaussian Color Model for skin segmentation," in *EUROCON, 2013 IEEE,* 2013, pp. 1653-1657.

[6] N. A. M. Isa, S. A. Salamah, and U. K. Ngah, "Adaptive fuzzy moving K-means clustering algorithm for image segmentation," *Consumer Electronics, IEEE Transactions on,* vol. 55, pp. 2145-2153, 2009.

[7] M. W. Ayech, K. El Kalti, and B. El Ayeb, "Image Segmentation Based on Adaptive Fuzzy-C-Means Clustering," in *Pattern Recognition (ICPR), 2010 20th International Conference on,* 2010, pp. 2306-2309.

[8] Y. Zhiding, Z. Ruobing, and Y. Simin, "A modified fuzzy c-means algorithm with adaptive spatial information for color image segmentation," in *Computational Intelligence for Image Processing, 2009. CIIP '09. IEEE Symposium on,* 2009, pp. 48-52.

[9] P. Hobson, B. C. Lovell, G. Percannella, M. Vento, and A. Wiliem, "Classifying Anti-nuclear Antibodies HEp-2 Images: A Benchmarking Platform," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, 2014, pp. 3233-3238.

[10] G. Iannello, L. Onofri, and P. Soda, "A bag of visual words approach for centromere and cytoplasmic staining pattern classification on HEp-2 images," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, 2012, pp. 1-6.

[11] A. B. L. Larsen, J. S. Vestergaard, and R. Larsen, "HEp-2 Cell Classification Using Shape Index Histograms With Donut-Shaped Spatial Pooling," *Medical Imaging, IEEE Transactions on*, vol. 33, pp. 1573-1580, 2014.

[12] S. Ghosh and V. Chaudhary, "Feature analysis for automatic classification of HEp-2 florescence patterns : Computer-Aided Diagnosis of Auto-immune diseases," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 174-177.

[13] V. Snell, W. Christmas, and J. Kittler, "Texture and shape in fluorescence pattern identification for auto-immune disease diagnosis," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 3750-3753.

[14] G. Thibault and J. Angulo, "Efficient statistical/morphological cell texture characterization and classification," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 2440-2443.

[15] P. Strandmark, J. Ulen, and F. Kahl, "HEp-2 staining pattern classification," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 33-36.

[16] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking HEp-2 Cells Classification Methods," *Medical Imaging, IEEE Transactions on,* vol. 32, pp. 1878-1889, 2013.

[17] C. Poynton, "A guided tour of color space," in *SMPTE Advanced Television and electronic imaging conference*, 1995, pp. 167-180.

[18] M. Sarifuddin and R. Missaoui, "A new perceptually uniform color space with associated color similarity measure for content-based image and video retrieval," in *Proc. of ACM SIGIR 2005 Workshop on Multimedia Information Retrieval (MMIR 2005),* 2005, pp. 1-8.

[19] V. Chernov, J. Alander, and V. Bochko, "Integer-based accurate conversion between RGB and HSV color spaces," *Computers & Electrical Engineering,* vol. 46, pp. 328-337, 2015.

[20] B. Fan, W. Fuchao, and H. Zhanyi, "Rotationally Invariant Descriptors Using Intensity Order Pooling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 2031-2045, 2012.

[21] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, 2004, pp. 1-2.