

**MITOTIC HEP-2 CELL RECOGNITION USING SUPPORT VECTOR
MACHINE UNDER CLASS SKEW**

By

SITI YASMIN BINTI SHEIKH ABDUL KARIM

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

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

© Copyright 2014
by
Siti Yasmin Sheikh Abdul Karim, 2014

CERTIFICATION OF APPROVAL

MITOTIC HEP-2 CELL RECOGNITION USING SUPPORT VECTOR MACHINE UNDER CLASS SKEW

By

Siti Yasmin Sheikh Abdul Karim
13392

A project dissertation submitted to the
Department of Electrical & Electronic Engineering
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
Bachelor of Engineering (Hons)
(Electrical & Electronic Engineering)

Approved:

Ms. Zazilah May
Project Supervisor

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK

January 2014

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 acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

Siti Yasmin Binti Sheikh Abdul Karim

ABSTRACT

A person with an autoimmune diseases will become hypersensitive to the surrounding that other normal person would usually not consider at all such as an allergy. This reaction happened when our immune system recognise our normal tissue as a dangerous foreign elements and proceed to attack them. The presence of antinuclear autoantibodies (ANA) in a patient serum can be detected by using the Indirect Immunofluorescence (IIF) image. By adding the mitotic cells into the well, the level of accuracy of the results achieved can be increased. The mitotic cells itself plays a crucial role in diagnosing an autoimmune diseases. This paper will focuses on the extracting the features of a mitotic HEP-2 cell in order to determine the presence of an ANA by noting the cells fluorescent-stained pattern, their intensity and also the presence of the mitotic cell itself. A skewed distribution of both mitotic and non-mitotic cells in the samples will also be considered to ensure the practicality of the project. To assist in the objectives, all the techniques used are explain in more detailed in this paper along with the result obtained by simulation from MATLAB for every steps from pre-processing to user interface menu. The procedures for the recognition of mitotic cells are image acquisition, pre-processing, segmentation, feature extraction and classification. The results obtained were tested using HEP-2 cell image datasets from MIVIA and from collaboration with Hospital Universiti Sains Malaysia (HUSM). The feature extractor used is the gray level co-occurrence matrix (GLCM) and classified using support vector machine (SVM) which will be presented in the RESULTS section.

ACKNOWLEDGEMENT

First and foremost, praise to God, the almighty for providing the opportunity and granting the strength to proceed and successfully completed this project and also with the involvement and contributions of certain parties, organizations and individuals. The author would like to take this opportunity to express the greatest gratitude to her supervisor, Ms Zazilah May for her exemplary guidance, monitoring and constant encouragement throughout the completion of this final year project. The blessing, help and guidance given by her from time to time shall carry me a long way in the journey of life on which I am about to embark.

Apart from that, special thanks as well to Dr. Josefina Barnachea Janier as the co-supervisor for her advice and patience guidance throughout the project which has indeed assists in the completion of this project and not to forget, Dr Ibrahim Faye for his kind guidance and teaching that has indeed contributed a lot in the achievement of this project. The author would like to extend her personal gratitude to all the authors' colleagues and family for all their help, cooperation and constant encouragement throughout the duration of this project.

The author appreciation also goes to University Teknologi PETRONAS (UTP) especially to Electrical & Electronics Engineering department and the FYP coordinators for all the seminars and constants guidance to all the final year students. Finally, a very deepest gratitude to all involved parties and individual for the help in both technical and non-technical aspects of this project which has in fact assisted me in the successful completion of this Final Year Project.

TABLE OF CONTENTS

ABSTRACT.....	III
ACKNOWLEDGEMENT	III
LIST OF TABLES	VII
LIST OF ABBREVIATIONS	VII
LIST OF FIGURES	VII
CHAPTER 1 PROJECT BACKGROUND	1
1.1 BACKGROUND STUDY.....	1
1.2 PROBLEM STATEMENT.....	2
1.3 OBJECTIVES AND SCOPE OF STUDY	2
1.4 RELEVANCY OF PROJECT	3
1.5 FEASIBILITY OF PROJECT.....	3
CHAPTER 2 LITERATURE REVIEW	4
2.1 MITOTIC CELLS	4
2.2 IMBALANCED DATA	6
2.4 SUPPORT VECTOR MACHINE.....	7
CHAPTER 3 METHODOLOGY	8
3.1 RESEARCH METHODOLOGY.....	8
3.2 PROJECT ACTIVITIES	9
3.3 GANTT CHART AND KEY MILESTONES	11
3.4 SOFTWARE REQUIRED.....	12

CHAPTER 4 RESULTS AND DISCUSSIONS.....	13
4.1 IMAGE ACQUISITION	13
4.2 PRE-PROCESSING AND SEGMENTATION.....	13
4.3 FEATURE EXTRACTION USING GLCM.....	16
4.4 CLASSIFICATION USING SUPPORT VECTOR MACHINE	18
4.5 RESULTS OBTAINED	20
CHAPTER 5 CONCLUSION AND RECOMMENDATION.....	23
5.1 CONCLUSION	23
5.2 RECOMMENDATION AND FUTURE WORK.....	24
REFERENCES.....	25
APPENDICES	27
APPENDIX A: CODING FOR PRE-PROCESSING AND SEGMENTATION	27
APPENDIX B: CODING FOR FEATURE EXTRACTION USING GLCM.....	28
APPENDIX C: CODING FOR CLASSIFICATION USING SVM	31
APPENDIX D: CODING FOR GUI MENU	32
APPENDIX E: CODING FOR TESTING A WHOLE IMAGE	33
APPENDIX F: CODING FOR TESTING A SINGLE IMAGE.....	35

LIST OF TABLES

Table 1: Examples of positive staining pattern	5
Table 2: Gantt Chart and Key Milestones for FYP 1.....	11
Table 3: Gantt Chart and Key Milestones for FYP 2.....	12
Table 4: SVM Class Distribution.....	18

LIST OF ABBREVIATIONS

ANA	Antinuclear autoantibodies
IIF	Indirect Immunofluorescence
HEP-2	Human Epithelial Type 2
HUSM	Hospital Universiti Sains Malaysia
CAD	Computer Aided Diagnosis
GLCM	Gray Level Co-occurrence Matrix
SVM	Support Vector Machine
GUI	Graphical User Interface

LIST OF FIGURES

Figure 1: Mitosis reproduction.....	4
Figure 2: Negative mitosis (A and B) and Positive Mitosis (C and D)	6
Figure 3: SVM Concept Illustration.....	7
Figure 4: Typical cell recognition modules	8
Figure 5: Project Activities	9
Figure 6: Flow Chart of the Project	10
Figure 7: HEp-2 Cell Image: Original Image (above) and Binary Image (bottom).....	14
Figure 8: HEp-2 Cell Image: Labelled image obtained using Label Matrix.....	15
Figure 9: 9 out of 85 Greyscale Cropped Image with Background Removed.	16
Figure 10: GLCM Cluster Prominence Feature Values.....	17
Figure 11: GLCM Cluster Shade Feature Values	17
Figure 12: GUI Menu Created	18
Figure 13: Result for Mitotic Cell.....	19
Figure 14: Result for Non-mitotic Cell.	19
Figure 15: Result Obtained with a Bounding Box.	20
Figure 16: Mitotic Cell recognition result on Centromere staining pattern.	21
Figure 17: Mitotic Cell recognition result on Homogenous staining pattern.....	21
Figure 18: Homogenous and Centromere Staining Pattern (from left to right).	22

CHAPTER 1

PROJECT BACKGROUND

1.1 Background Study

Our immune system keeps our organism healthy by recognizing and defending us from foreign elements. These foreign elements are called pathogens and they are infectious elements that attack us, i.e. bacteria and viruses. The immune system makes a plenty of proteins called antibodies and antibodies has indeed play a key role in our immune system since they are capable of detecting and attacking those infectious elements in the body.

However, sometimes these antibodies make a mistake, where they identify normal, naturally-occurring proteins in our bodies as foreign and dangerous instead. As a result, they attack the person's own tissues. This reaction is known as antinuclear autoantibody (ANA) [1]. When the immune system can no longer differentiate between normal body tissues and pathogens, a reaction will occurred that we classify as an autoimmune disease. The immune system will become hypersensitive and response to normal body tissue that it normally would have ignored.

Detection of ANA is the basic and most crucial in detecting an autoimmune diseases, and Indirect Immunofluorescence (IIF) staining method has been the preferred method in diagnosing the presence of ANA from a patient serum [2, 3]. IIF image with fluorescent-stained mitotic HEp-2 cells is very important to detect the presence of ANA in those diseases. The IIF slides were examined under a fluorescence microscope, and all important details were noted which are the fluorescent staining pattern, the fluorescence intensity, and also the presence of a mitotic cells. Looking for the mitotic cells is in fact an essential element of the diagnostic procedure.

1.2 Problem Statement

IIF image with HEp-2 cells is used for the detection of ANA in autoimmune diseases; however, since IIF image analysis is subjected to subjective interpretation and also by the differences in the HEp-2 substrates, it will give rise to inaccuracy and classification variability. The classifying process will highly depend on the basis of one skills and experience [3, 4]. Therefore, the presence of a mitotic cell during the process verifies correctness of the slide preparation process and also provides information on the fluorescent-stained image pattern. Alas, works made on the methods suitable to classify such cells were very limited. Most recent research has been instead directed to IIF image acquisition, image segmentation, fluorescence intensity classification and staining pattern recognition [2].

Traditional recognition algorithms cannot be employed to successfully identify the mitotic cells as they are more skewed towards the majority class, resulting in low accuracy towards the minority class [5]. In a set of cells samples, the amount of mitotic cells and the non-mitotic cells differs and thus a skewed nature of the recognition needs to be considered. Both the minority and the majority classes needed to be identified equally without biasing towards either class to ensure a high accuracy results.

1.3 Objectives and Scope of Study

The main objective of this project is:

- To identify the features of mitotic cells.
- To develop recognition algorithm for mitotic cells using Support Vector Machine under class skew.
- To validate and optimize result obtained

This project will focus on the presence and features of a mitotic cell by using the best feature extraction algorithm available that is the most suitable for extracting the features of a mitotic HEp-2 cells and classified using SVM. The technique chosen must also be feasible to be completed within the time range given for the project (2 semesters).

1.4 Relevancy of Project

The relevance of this project is because over the last few years, the growing on the incidence of autoimmune diseases was observed and a higher number of health structures laboratories are needed to encourage awareness on this issue and most importantly to discover an improvement and introduce new techniques for diagnosing [6].

For that reason, by implementing Support Vector Machine (SVM), it is expected to produce more accurate results than the existing technique for mitotic Hep-2 cell recognition which depends on skills of the one doing the analysis [2].

1.5 Feasibility of Project

This project is feasible enough to be done within two (2) semesters, as it mostly only involves simulation using software. However, clear understanding on the Support Vector Machine (SVM) and its algorithm will indeed result in faster outcome obtained at the end of the project.

CHAPTER 2

LITERATURE REVIEW

2.1 Mitotic Cells

Mitosis is a process of nuclear division in eukaryotic cells that occurs when a parent cell divides to produce two identical daughter cells [7]. Figure 1 shows mitosis form of reproduction. In the process of detecting autoimmune diseases, the presences of mitotic cells have been proved to be beneficial.

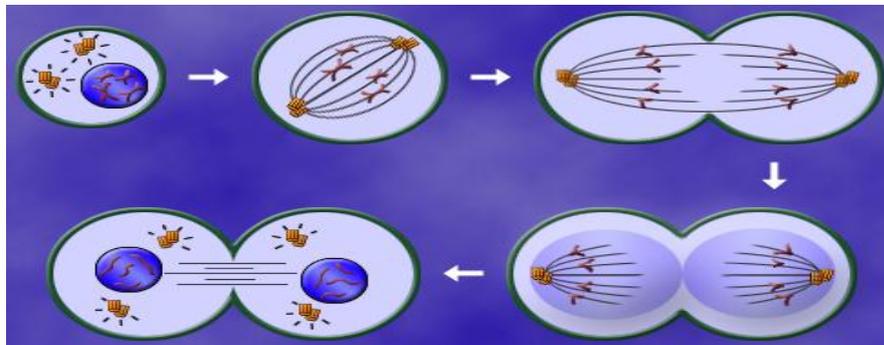


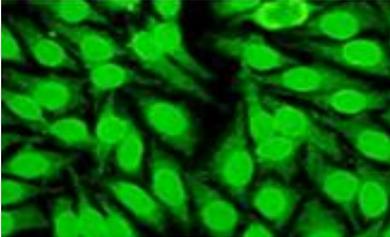
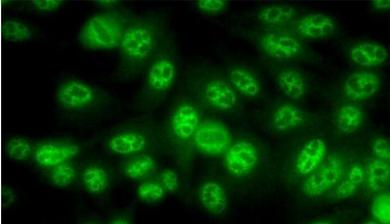
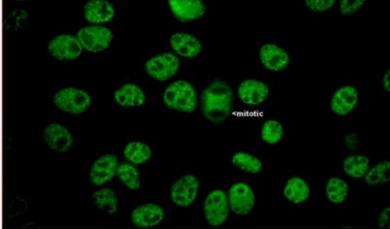
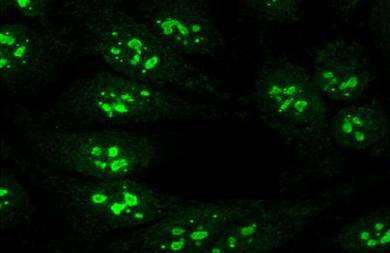
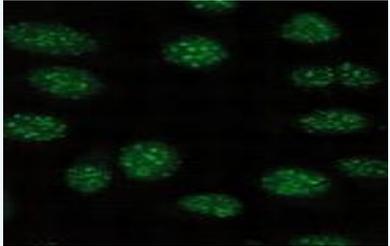
Figure 1: Mitosis reproduction

Detection of ANA in autoimmune diseases using the IIF image with HEp-2 cells, is when specific fluorescence pattern can be observed in the serum on the humane epithelial cell line (HEp-2) [8]. Current guideline for ANA tests recommends the use of HEp-2 substrate, with a serum dilution of 1:80 and requires classifying both fluorescence intensity and staining pattern [2, 9]. Since IIF is a quite biased, the diagnosis is usually performed by highly trained personal [1].

The two (2) main importance of adding mitotic cells to slides are to ensure that the well has been prepared properly and because mitotic cell provide information on the staining pattern. By detecting at least one fluorescent mitotic cell, medical officers can be assured that the preparation process were done correctly. The mitotic cell itself has its own uniqueness in terms of their staining pattern which can provide information caused by the mitosis cycle [2].

Positive HEp-2 sample may have many identified positive patterns however in this paper, we will consider only 2 of the positive HEp-2 pattern which are the Centromere pattern and the homogenous pattern [10]. Table 1 list down the 5 autoantibody fluorescence pattern.

Table 1: Examples of positive staining pattern [2]

Patterns		Description
Homogenous		A diffuse fairly uniform staining across the whole nucleus.
Fine Speckled		A fine or coarse granular nuclear staining of the interphase cell nuclei [6].
Coarse Speckled		
Nucleolar		A small number (less than 6) of larger bright areas within the nucleus.
Centromere		Large numbers of strong bright spots on a darker background. Several discrete speckles around 40–60 are present distributed throughout the interphase nuclei.

Mitotic cells have some peculiarities pattern because of the mitosis cycle and its substrate of HEp-2 which exhibit two fluorescent patterns that are shown below in Figure 2. Column C and D represent the positive mitosis where the collapsed chromosomes mass located in the middle part of the cell is darker or has a weak fluorescence pattern while the outer region has high fluorescence intensity. Meanwhile the columns A and B represent the negative mitosis which is the vice versa of the positive mitosis; the cell body is does not has a weak fluorescence, while the chromosomes mass in the middle has high fluorescent intensity.

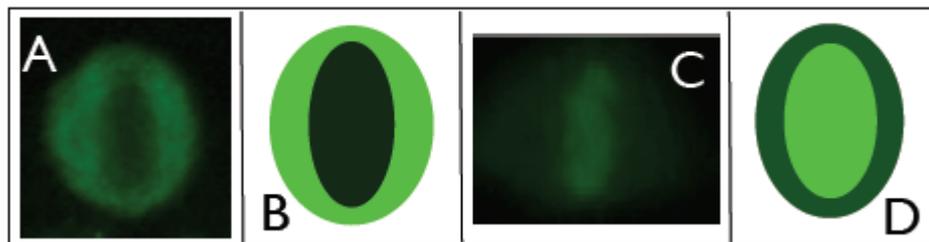


Figure 2: Negative mitosis (A and B) and Positive Mitosis (C and D) [2]

It is worth observing that collapsed chromosomes mass has a circular or elliptic shape, regardless of the six staining pattern from Table 1. [2]

2.2 Imbalanced Data

Imbalanced data is a set of data which has a high degree of imbalanced that the data can be classified into either minority class or majority class. Classifying and identifying the minority class has been a challenge and are extensively researched by researcher [11]. When the a priori sample distribution is skewed, traditional classification algorithm are not suitable as they are designed to minimize error over training samples, ignoring classes composed of few instances. As a result, poor predictive accuracy towards the minority one will occur. The class skew is when a sample consisted of an imbalanced data of classes. Researches are usually made using an artificially balanced sample of mitotic and non-mitotic cell with 50:50 ratios. This is in fact not practical in real life situations.

An example of an imbalanced data that occurs in real life are fraud detection [12], detecting a rare diseases [13], indentifying specific users of telecommunication company [14] and many more. However unlike cases of credit card fraud or text classification, failing to detect the minority instances for medical cases may cause

serious consequences. For instance, only 1% of the blood serum contains a disease cell. In such situation, labelling all the blood samples as clear from diseases would give an accuracy of 99%, but failing on the entire negative sample. This is indeed very dangerous in real life cases [11, 15].

2.3 Gray Level Co-occurrence Matrix (GLCM)

Gray level co-occurrence matrix (GLCM) is one of the methods of extracting from images. There are many features that can be extracted from GLCM [18]. GLCM is a second-order estimation based on the probability that two pixels with grey levels occur for a given distance and direction. A matrix of dimensions equals to the gray level of the image is produced for each distance and orientation [16]. There are a total of 22 features that can be computed by GLCM such as Autocorrelation, Contrast, Correlation, Cluster prominence, Cluster shade, Dissimilarity, Energy, Entropy, Homogeneity, etc but only the features that show a distinct values between the two classes are selected for the recognition of a mitotic cells.

2.4 Support Vector Machine

Support Vector Machine (SVM) is one of the most popular machine language tool that have been use as a classifier in pattern recognition and data analysing. SVM are based on the concept of decision planes [17] and this plane separates the classes by creating a clear boundaries between them. Figure 3 below is the illustration to explain the concept of a simple two class SVM.

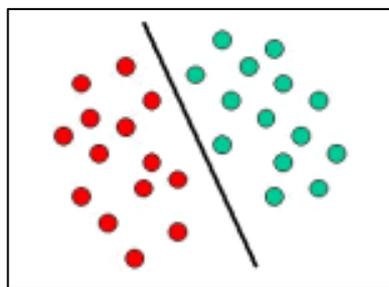


Figure 3: SVM Concept Illustration [17]

A linear SVM is used in this project for the recognition of two classes which is mitotic class and the non-mitotic class. Each class has their own specific features value ranges which are distinctive to one another. Consequently, for this project, the new unknown class will either fall into the mitotic or the non-mitotic classes only.

CHAPTER 3

METHODOLOGY

3.1 Research Methodology

A typical cell recognition process consists of 5 steps that needed to be follow to ensure the results obtained are accurate. The steps include image acquisition, pre-processing, segmentation, feature extraction, and lastly classification and recognition. The schematic diagram of the proposed recognition system is shown in Figure 4.



Figure 4: Typical cell recognition modules

Even though, this paper focuses on the features extraction and classification of the mitotic HEp-2 cells; all the steps from the cell recognition modules must be completed step-by step. For the other steps apart from classification, we will choose the most suitable and simple method available as the main focuses in this project was to extract the features of a mitotic HEp-2 cell. The second and third step of the whole process which is the pre-processing and segmentation has been done and tested on every HEp-2 cell image use for training and testing.

For the pre-processing, the image is filtered out from noise, converting into greyscale and binary image while the segmentation steps is to compute the borderline of each cell and labelled each connected components. Gray level co-occurrence matrix (GLCM) is used as the feature extractor and as per title; support vector machine (SVM) will be use as the classifier.

3.2 Project Activities

Figure 5 shows the flow of the project activities for both FYP 1 and FYP 2 and the flow chart of the project which is the detailed process flow are presented in Figure 6:

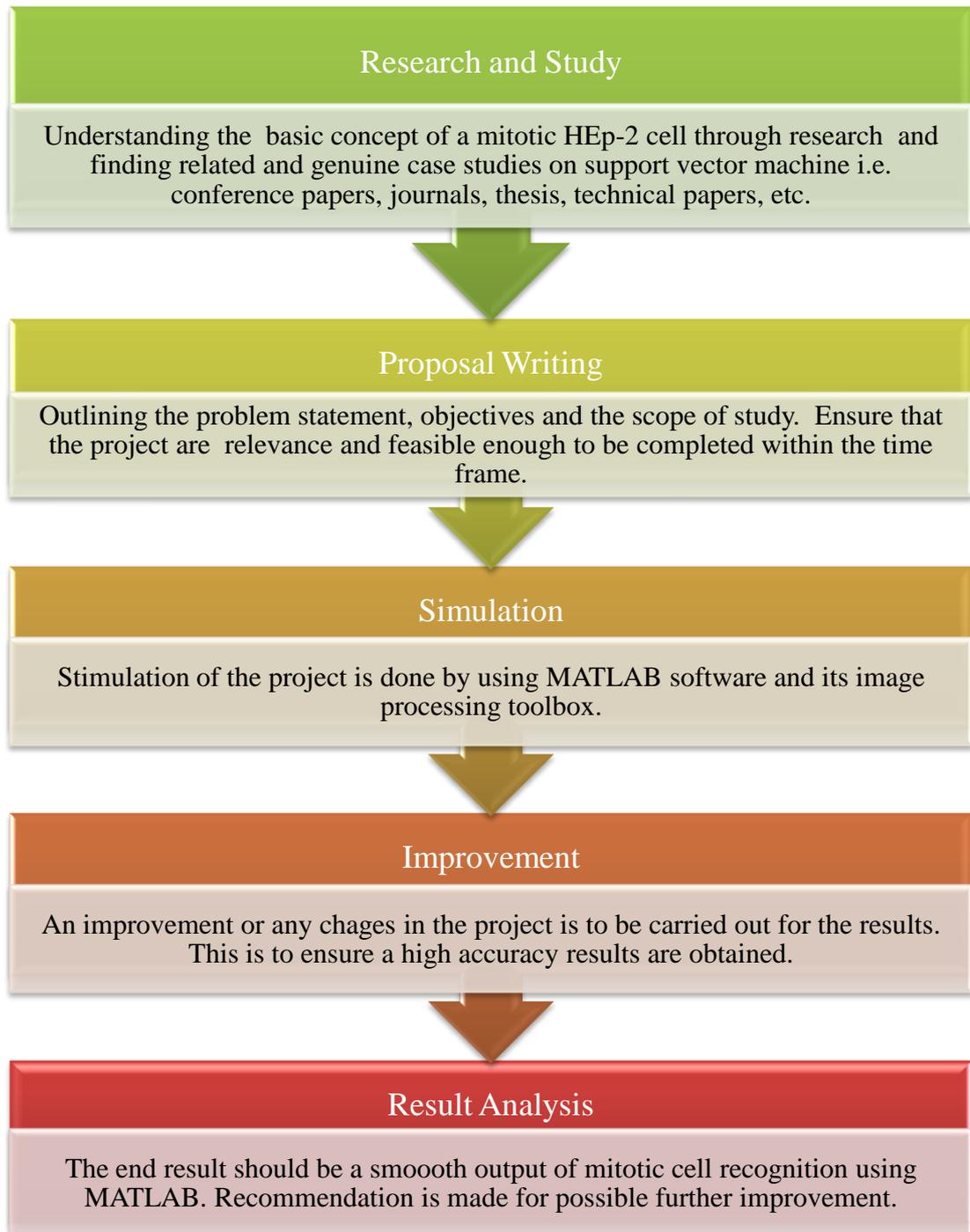


Figure 5: Project Activities

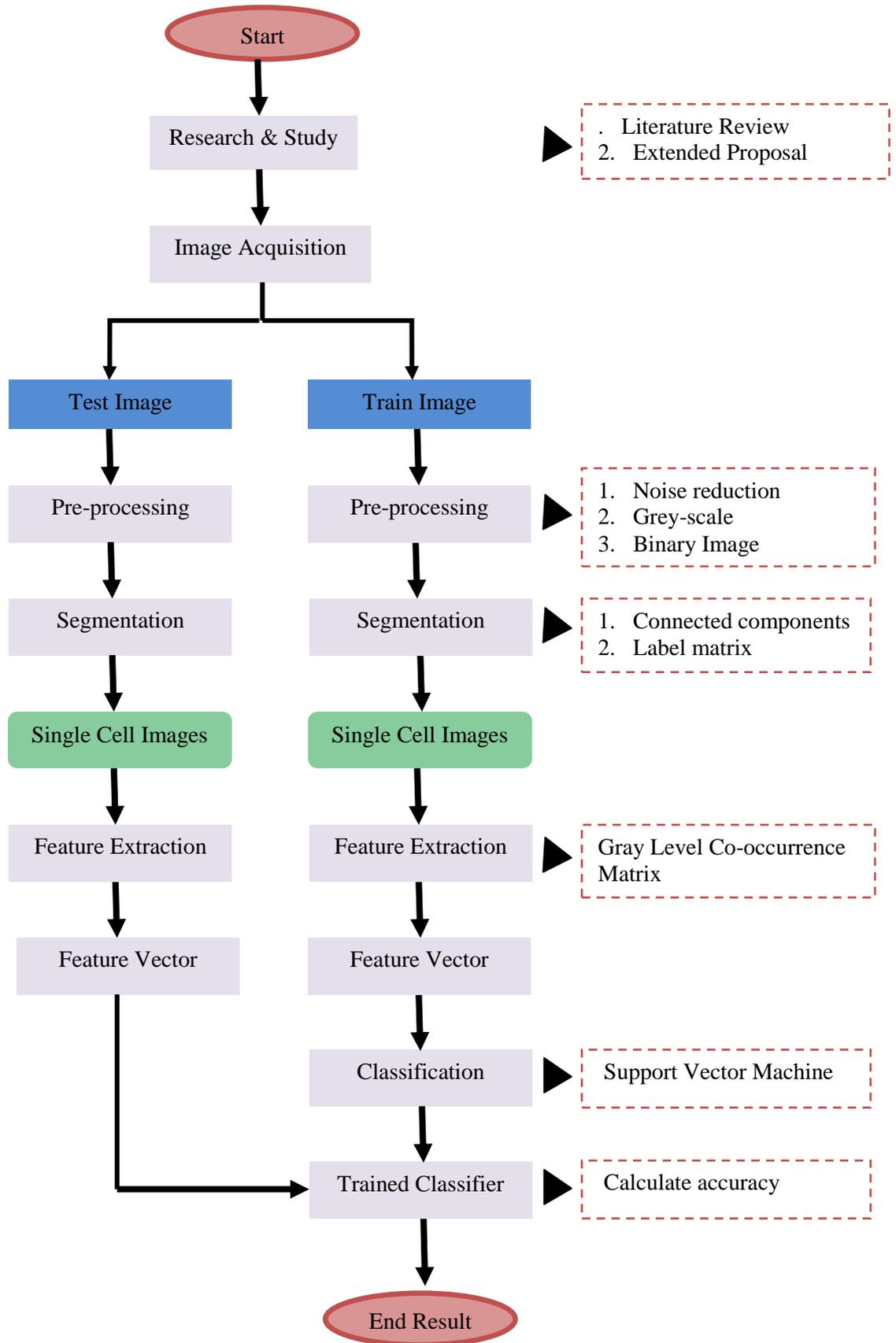


Figure 6: Flow Chart of the Project

3.3 Gantt Chart and Key Milestones

The Table 2 and 3 below explained the key milestone and the Gantt Chart for both FYP 1 and FYP 2.

Table 2: Gantt Chart and Key Milestones for FYP 1

Details	Week													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Title Selection and Confirmation	■	■												
2. Literature Review Research			■	■	■	■								
3. Preparation of Extended Proposal					■	■								
4. Extended Proposal Submission						●								
5. Preparation for Proposal Defence						■	■	■	■					
6. Proposal Defence and Progress Evaluation									●					
7. Train Image - Pre-processing - Segmentation										■	■	■	●	
8. Preparation of Interim Report												■	■	
9. Draft of Interim Report Submission													●	
10. Improvement on Interim Report													■	■
11. Interim Report Submission														●



Process



Key Milestone

Table 3: Gantt Chart and Key Milestones for FYP 2

Details	Week														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Train Image Segmentation	■	■													
2. Feature Extraction: Gray Level Co-occurrence Matrix			■	■	■	■	■	■							
4. Compute Feature Vector								■							
3. Progress Report Submission								●							
5. Train Classifier : Support Vector Machine (SVM)								■	■						
6. Testing Image using trained SVM Classifier									■	■	■				
7. Calculate Accuracy											■				
6. Pre-EDX/SEDEX											●				
7. Preparation of Draft Final Report											■	■	■		
8. Draft Final Report submission													●		
9. Improvement on Final Report													■	■	
10. Final Report Submission														●	
11. Viva															●



Process



Key Milestone

3.4 Software Required

Apart from the Microsoft Office used throughout the period of completing this project, the MATLAB software using the image processing toolbox will be the main tool use to simulate all the result and undergo all the cell recognition process and also the MATLAB GUI.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Image Acquisition

The dataset acquire are from the HUSM and MIVIA HEp-2 images dataset from a research project “Classification of Immunofluorescence Images for the Diagnosis of Autoimmune Diseases”, supported by “Regione Campania, Italy”. It consists of images obtained with a fixed dilution of 1:80 HEp-2 substrates as recommended by guidelines. The images were acquired using a fluorescence microscope (40-fold magnification) coupled with a 50W mercury vapour lamp and a digital camera. The camera has a CCD with squared pixel of equal side to 6.45 μm . The images have a resolution of 1388 \times 1038 pixels, a colour depth of 24 bits and they are stored as bitmap images.

The dataset composed of 28 images organized with respect to the 6 main patterns that consist of 1527 cells, with 70 mitotic cells and 1457 non-mitotic cells. This shows a strong degree of imbalanced dataset with the minority class of the mitotic cell being 4.58% out of the total dataset. The non-mitotic cells display one of the six main staining patterns, however, only patterns of centromere and homogenous will be considered in this project. Images used are divided into training images and testing images. As shown from Figure 6, each testing and training images will undergo the same processes from pre-processing to feature extraction.

4.2 Pre-processing and Segmentation

After acquiring the image, the first step is the pre-processing where the noise is filtered out and the image is converted into greyscale and binary. Converting the image into greyscale is very important in order to use GLCM as a feature extractor and also serve as the first step before converting the image into binary images. Only cells that are captured as a white mask will be considered because the black areas cover a considerable area of the image and could affect the extracted features. Therefore, each white area on the mask or connected components, are labelled and

assigned a unique number using labelmatrix function. From the labelled cell obtained, each cell from the original green image is crop to have an image of each cell with the background removed. For training, an image of centromere patterned HEp-2 cell is used. The image is converted into binary image; as shown in Figure 7, before it can be labelled and cropped.

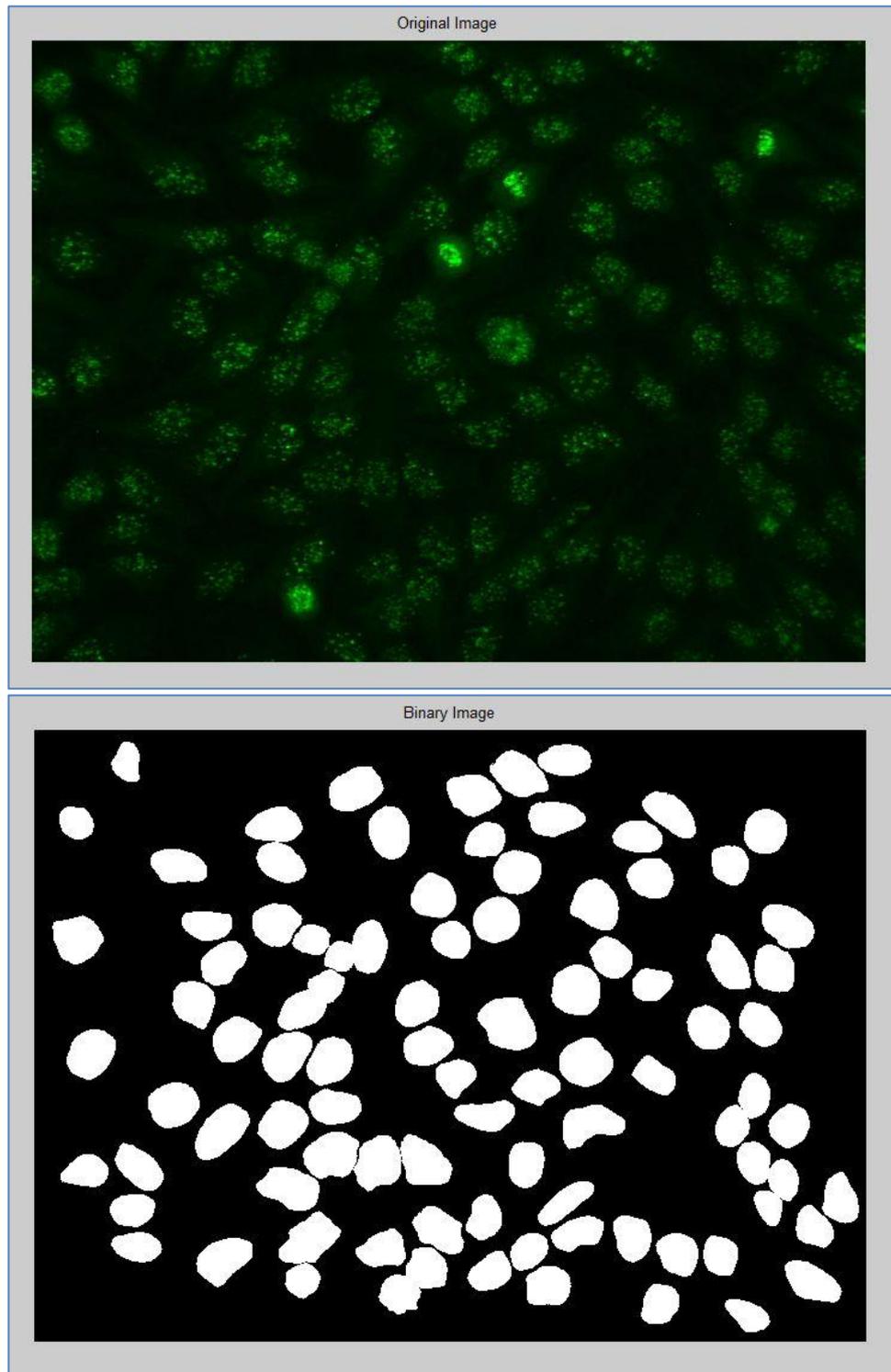


Figure 7: HEp-2 Cell Image: Original Image (above) and Binary Image (bottom)

A total of 85 cells are detected and labelled as shown on Figure 8. The labelling of an image uses the label matrix function, which is based on locating the connected components (from the binary image obtained). Therefore, there will be more than one cell that are labelled as one because some cells are too close to each other that their pixels are joined (e.g. cell number 27 from Figure 8).

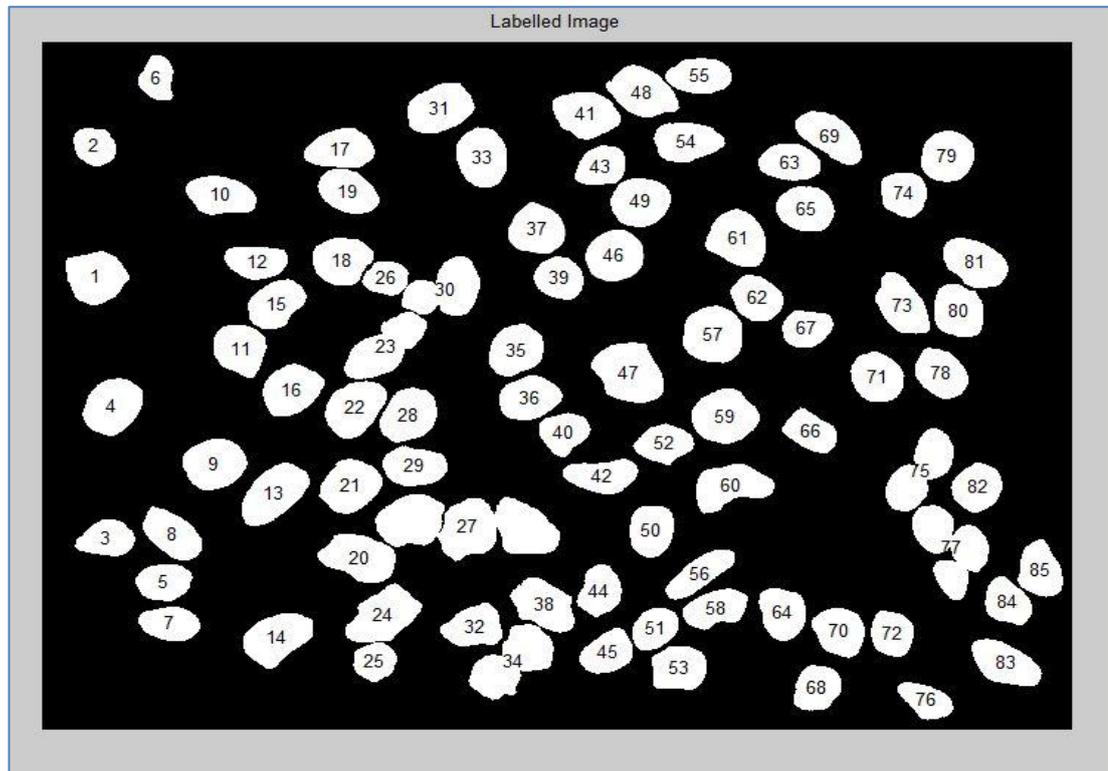


Figure 8: HEp-2 Cell Image: Labelled image obtained using Label Matrix.

From the number assigned previously, all 85 cells are then cropped and imposed on the original image. Each image will have a single cell with all the background remove in order for the particular cell features to be extracted. The Figure 9 shown shows the 9 out of 85 cell image obtained that has been converted to greyscale. Each cell image is automatically converted into greyscale, as GLCM features extraction algorithm requires grey image in order to obtain its feature vector. For training images, the feature vector obtained are used to train the classifier while for testing, the feature values obtained are classified using the trained classifier.



Figure 9: 9 out of 85 Greyscale Cropped Image with Background Removed.

4.3 Feature Extraction using GLCM

After the cells are segmented using label matrix, the features of each cell are extracted to obtain their feature values. Features extracted by GLCM will have a feature vector and that vector can be considered as a sample for training data. From a single image of a centromere HEp-2 cells, we have feature vectors for all 4 mitotic cell and 81 non-mitotic cells. Those feature vectors are then used to train a machine like SVM to recognize the type of cells.

Gray Level Co-occurrence algorithm is used as the main feature extractor. As GLCM has a total of 22 features, only the most suitable feature that can fully capture the features of a mitotic cell is selected. Blobs that have maximum green values are located to identify the mitotic cells, and after testing an image of mitotic and non-mitotic cells, the most prominent feature that can be captured from GLCM are the cluster prominence and cluster shade as shown in Figure 10 and 11.

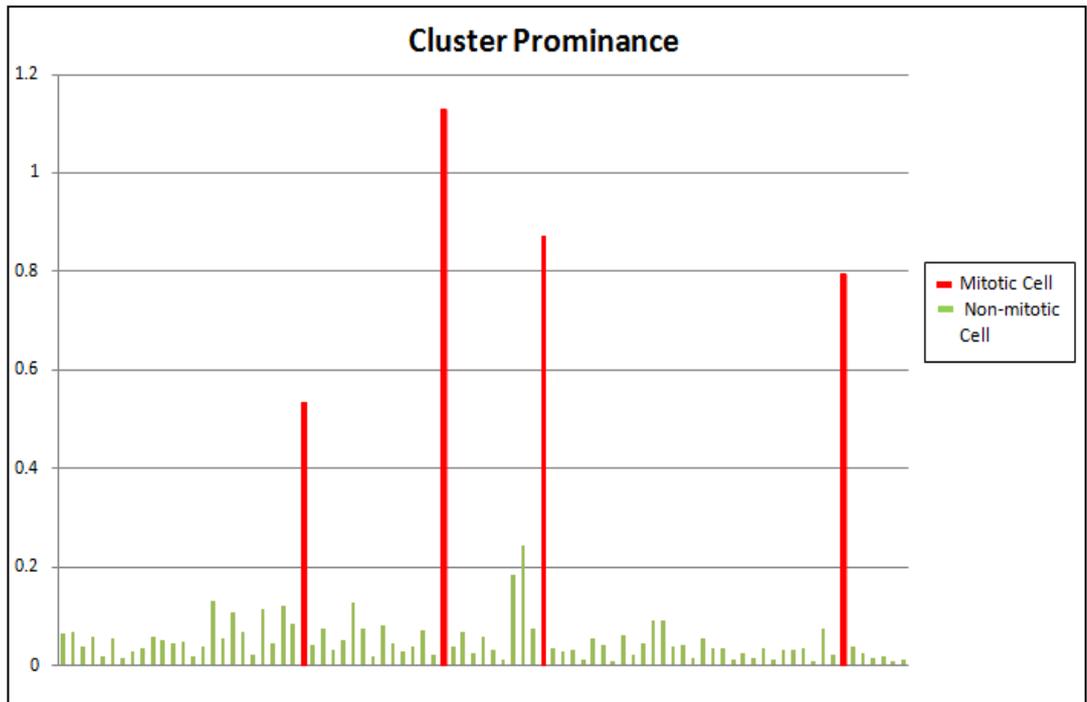


Figure 10: GLCM Cluster Prominence Feature Values

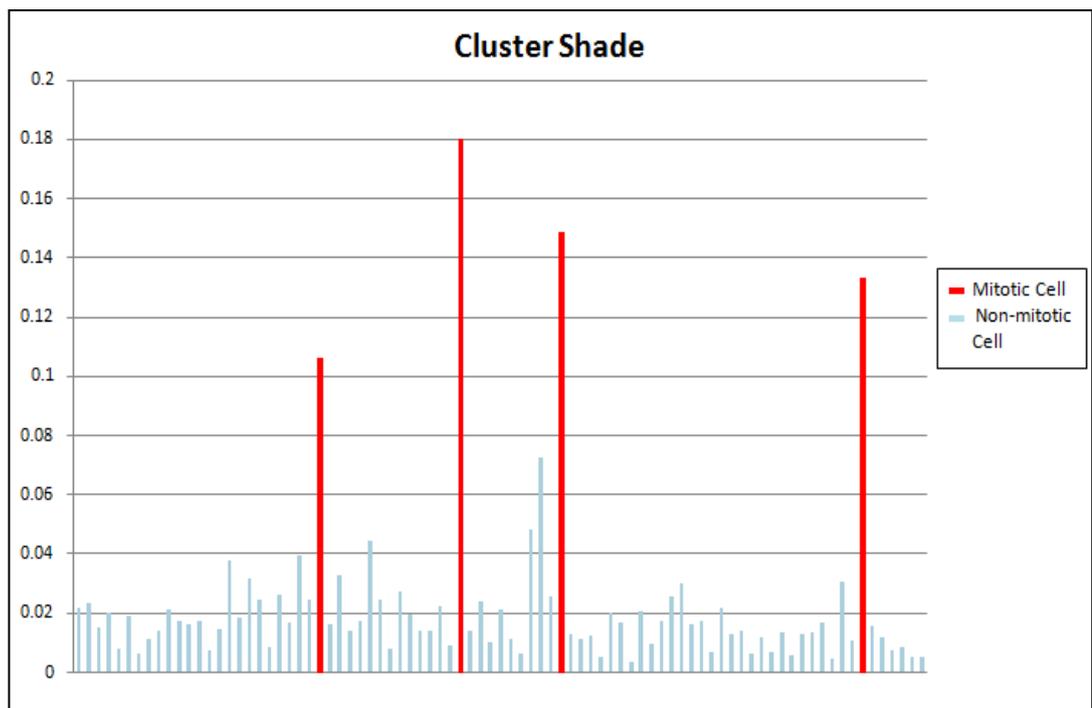


Figure 11: GLCM Cluster Shade Feature Values

4.4 Classification using Support Vector Machine

From the feature values obtained by using GLCM, each value are categorized according to their classes which are then used to train the SVM as shown in the Table 4 below:

Table 4: SVM Class Distribution

	Feature Value Ranges	Class Type
Non-mitotic Cell	0-0.2	Class 0
Mitotic Cell	0.21>	Class 1

All the testing images will be tested using the trained classifier and results obtained should either be in the Class 0 or Class 1. In order for the project to be user friendly, a simple GUI menu was created where user can choose whether to test a single image of a cell or to test a whole image of a cell. Figure 12 below shows the GUI menu created with a simple selection button.

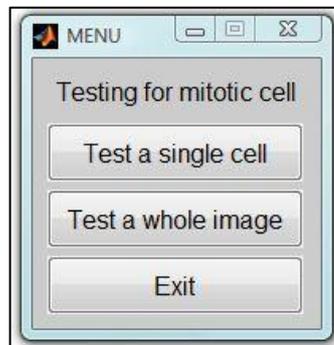


Figure 12: GUI Menu Created

When the user select the first option, “Test a single cell”, the program will allow users to select an image of a single cell for testing and a popup window will appear with the result of the testing indicating whether it is a mitotic cell or not a mitotic cell. Figure 13 shows the result when testing for a mitotic cell while Figure 14 shows the result when testing a non-mitotic cell.

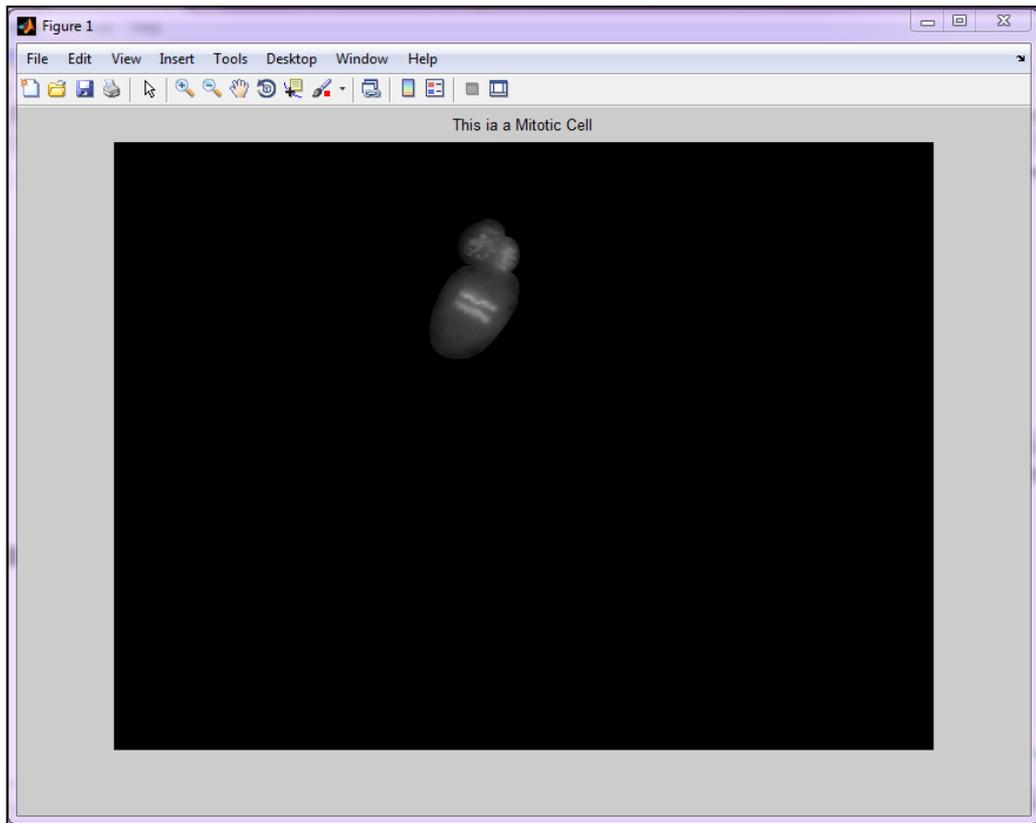


Figure 13: Result for Mitotic Cell.

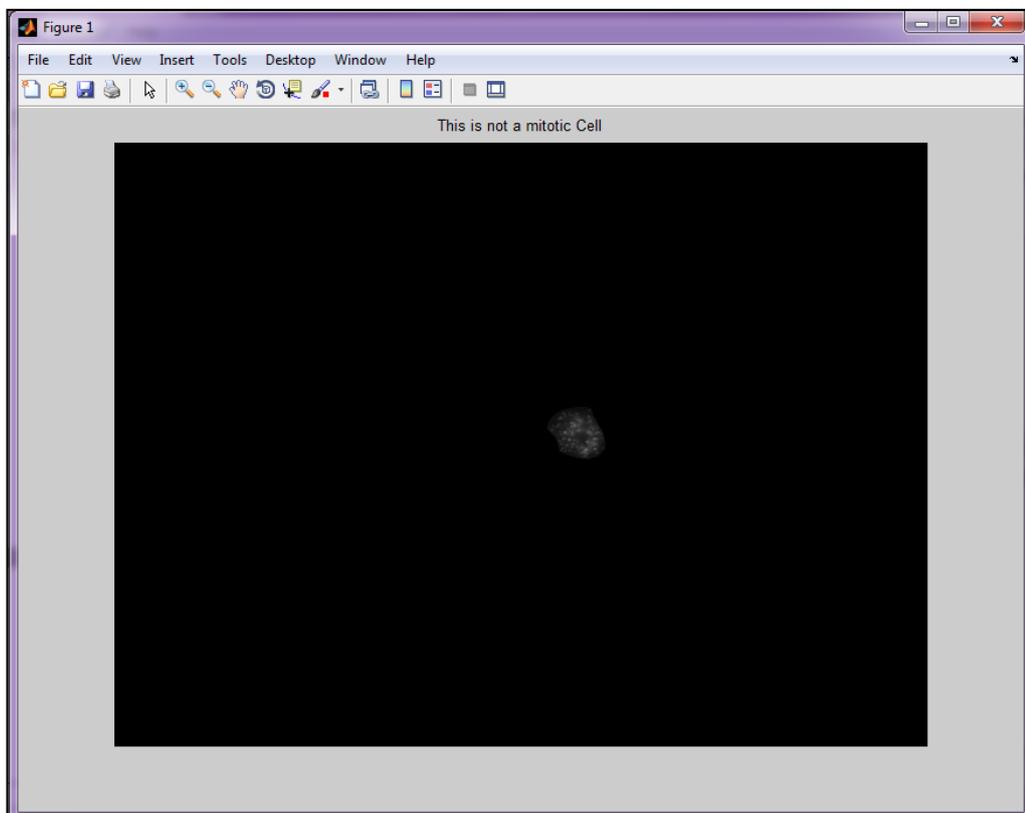


Figure 14: Result for Non-mitotic Cell.

When users select the second option “Testing a whole image”, a popup menu will appear for the user to select the desired cells image for testing. The program will process the image and create a bounding box around the cells that are detected as mitotic as shown in Figure 15 below:

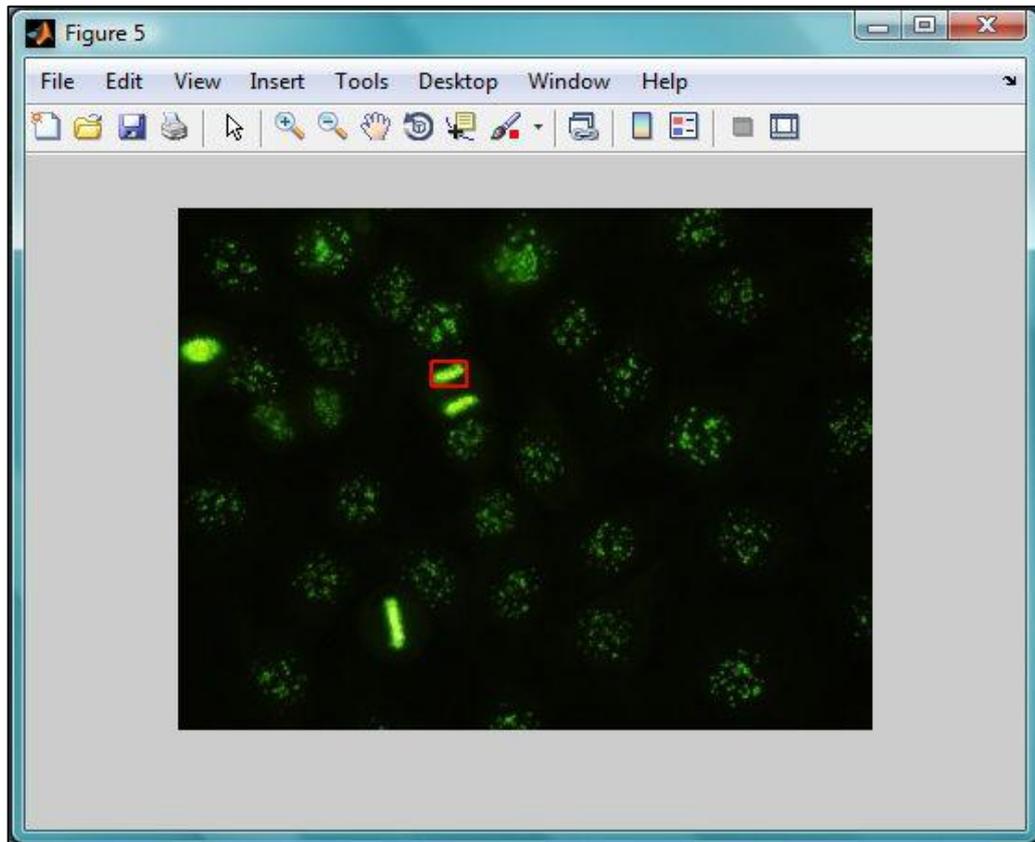


Figure 15: Result Obtained with a Bounding Box.

4.5 Results Obtained

Two different types of pattern which is centromere and homogenous where tested and the result of the classification accuracy when tested on a whole cell image are calculated in Figure 16 and Figure 17.

A total of 4 images for each pattern are tested and the numbers of mitotic cell detected by the classifier are calculated. For centromere stained pattern image 1, there are a total 4 mitotic cell but instead the machine classified about 7 mitotic cells which have an additional of 3 misclassified cells. The second image has 100% accuracy, while the 3rd and 4th image has a ± 2 misclassify cell respectively.

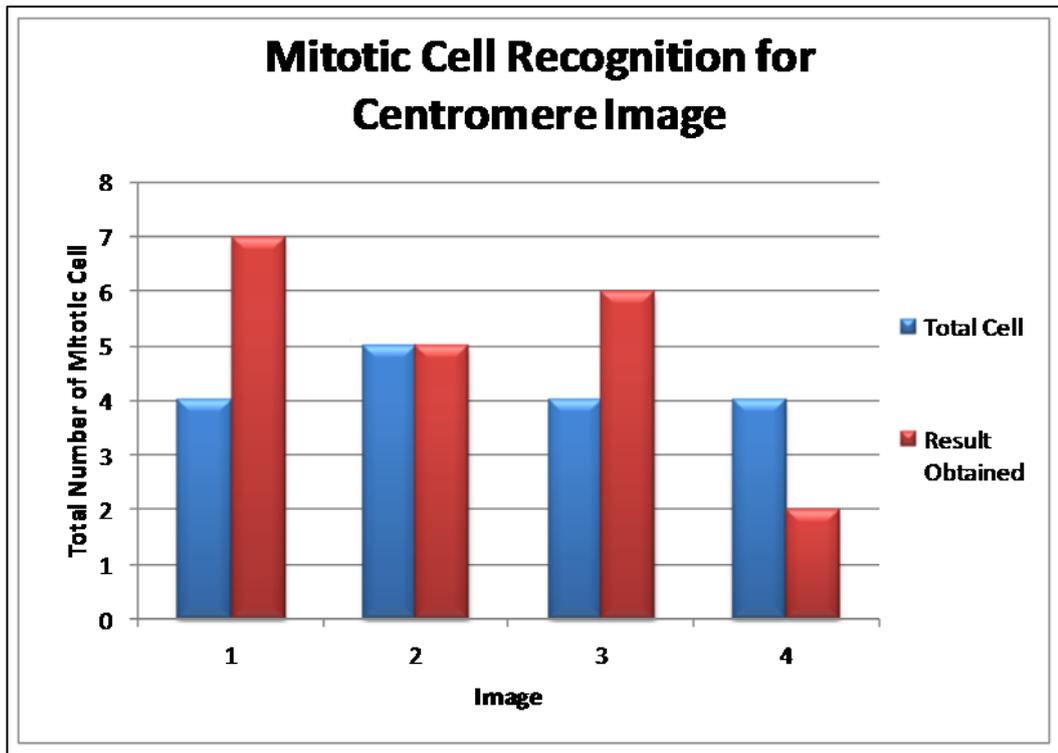


Figure 16: Mitotic Cell recognition result on Centromere staining pattern.

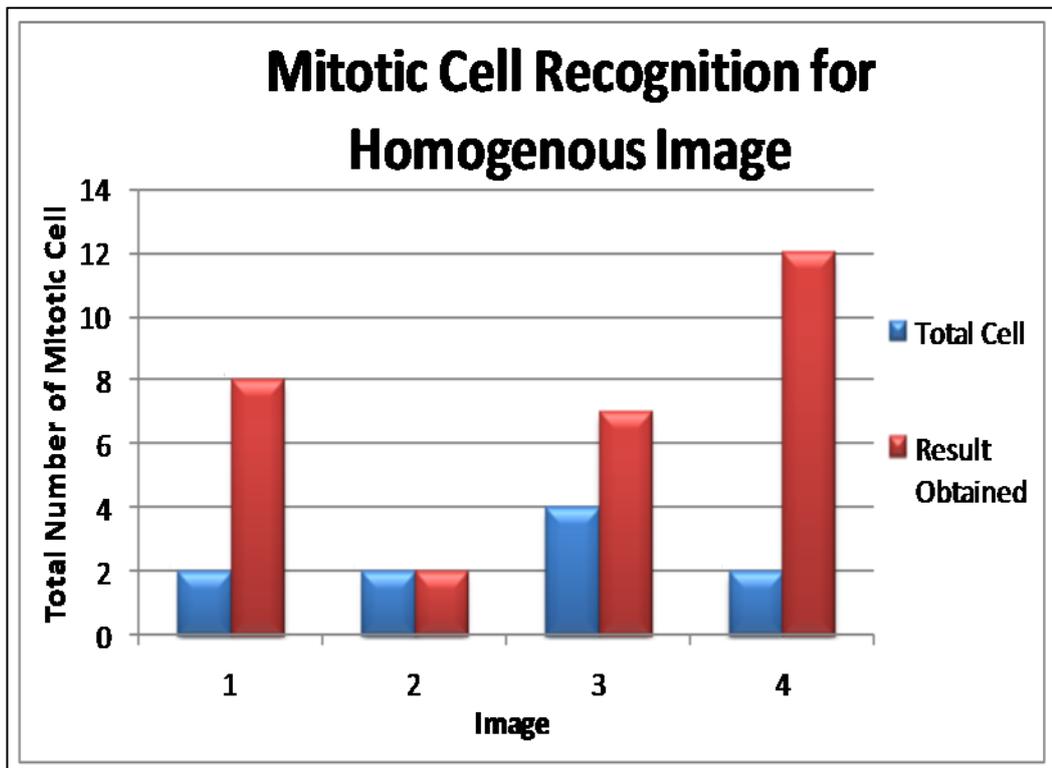


Figure 17: Mitotic Cell recognition result on Homogenous staining pattern.

However, the result obtained for the homogenous pattern shows a poorer accuracy compared to the centromere pattern. The highest numbers of misclassified cells are from image number four with the total mitotic cells to be four but instead the classifier recognises an additional of 12 cells.

This is due to the fact that the staining pattern for homogenous is slightly similar to a mitotic cell compared to the centromere staining pattern, hence it is reasonable for the classifier to misclassify some cells as mitotic cells instead as shown in figure below.

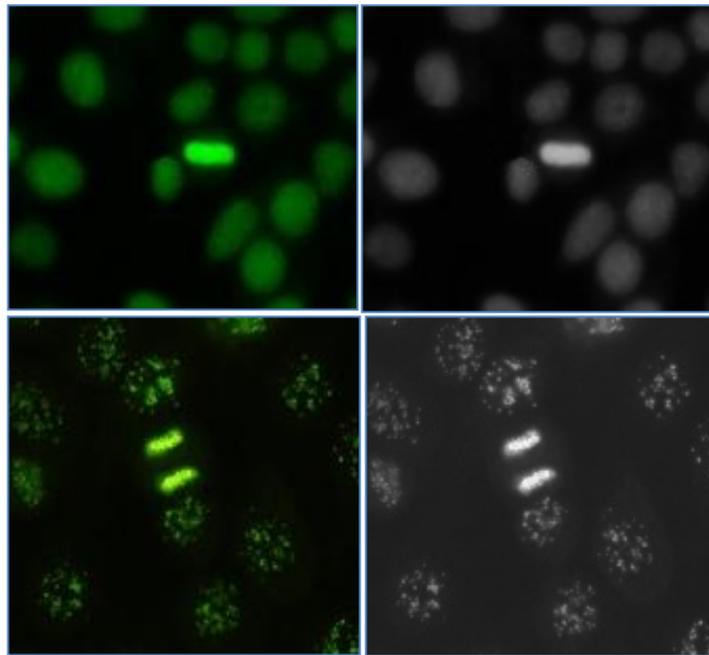


Figure 18: Homogenous (above) and Centromere Staining Pattern (bottom).

CHAPTER 5

CONCLUSION AND RECOMMENDTION

5.1 CONCLUSION

In this paper, we have presented the general concept of autoimmune diseases and the steps in diagnosing after using the IIF method which is one the most preferred method in detecting the presence of antinuclear autoantibody (ANA). Discovering autoimmune diseases are in fact very crucial as this disease involves our antibody to attack our normal body tissue and might cause disorder such as Addison's disease, pernicious anaemia, reactive arthritis, type I diabetes, multiple sclerosis and many more.

The use of mitotic cells in the diagnostic process is very important as it contributes to the increase in accuracy and also ensures the correctness of the preparation. Detecting the mitotic cells using the traditional approach are not suitable as they only considered a skewed distribution of the samples and may tend to be bias towards the majority class (non-mitotic cell) and ignoring the minority class (mitotic cell). This may be acceptable in real life cases such as fraud but is totally unacceptable in medical cases.

The steps in cell recognition involve image acquisition, pre-processing, segmentation, feature extraction and lastly classification where gray level co-occurrence matrix (GLCM) is use as the main feature extractor while the support vector machine (SVM) as the classifier. Based on the result obtained, the combination of GLCM and SVM are indeed reliable in detecting the mitotic cells and for skewed data.

5.2 RECOMMENDATION AND FUTURE WORK

Two main problems which occur throughout the studies would be that only image with a positive intensity can be classify accordingly while image with intermediate intensity tends to classify towards the majority classes or non-mitotic cells. This shows a perfect example of the effect on an imbalanced data on a classifier. An improvement on the future would be to classify image on other intensity level as well by adding more feature extractor and also the classifier. This is to increase the accuracy of the overall performance as each feature extractor and a machine language tool has their on specialty.

Adding more feature extractor may also assists in the improvement of classifying the mitotic cells for homogenous staining pattern. This is because the homogenous staining pattern is slightly similar to the features of the mitotic cells as shown on figure 18, hence the misclassification.

Future work on this project can also be focuses on other four staining patterns as during the studies only two staining patterns are considered which is the homogenous and centromere. As each staining pattern represent a specific diseases, being able to classify all the staining pattern will be a great advantages on the future of autoimmune diseases detection using computer aided diagnosis (CAD).

REFERENCES

- [1] 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*, Tsukuba, Japan, 2012, pp. 3750-3753.
- [2] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Early experiences in mitotic cells recognition on HEp-2 slides," in *Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on*, Perth, Australia, 2010, pp. 38-43.
- [3] G. Percannella, P. Soda, and M. Vento, "Mitotic HEp-2 Cells Recognition under Class Skew," in *Image Analysis and Processing – ICIAP 2011*. vol. 6979, G. Maino and G. Foresti, Eds., ed: Springer Berlin Heidelberg, 2011, pp. 353-362.
- [4] S. Di Cataldo, A. Bottino, E. Ficarra, and E. Macii, "Applying textural features to the classification of HEp-2 cell patterns in IIF images," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, Tsukuba, Japan, 2012, pp. 3349-3352.
- [5] J. S. S. R. Barandelaar, V. Garcia, E. Rangel, "Strategies for learning in class imbalance problems," in *Pattern Recognition* vol. 36, ed: Elsevier Science Ltd., 2002, pp. 849 – 851.
- [6] G. I. Paolo Soda, "SlideImaging: classification of IIF images," Integrated Research Centre, University Campus Bio-Medico of Rome, Rome, Italy, 2009.
- [7] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*: Springer Publishing Company, Incorporated, 2007.
- [8] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, "HEp-2 Cells classification via fusion of morphological and textural features," in *Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on*, Larnaca, Cyprus, 2012, pp. 689-694.
- [9] H. Tsu-Yi, H. Yi-Chu, C. Chia-Wei, and H. Yu-Len, "HEp-2 cell classification in indirect immunofluorescence images," in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, Beijing, China, 2009, pp. 1-4.
- [10] S. Ghosh and V. Chaudhary, "Feature analysis for automatic classification of HEp-2 fluorescence patterns : Computer-Aided Diagnosis of Auto-immune diseases," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, Tsukuba, Japan, 2012, pp. 174-177.
- [11] A. Aşkan and S. Sayın, "SVM classification for imbalanced data sets using a multiobjective optimization framework," *Annals of Operations Research*, pp. 1-13, 2013/01/15 2013.
- [12] T. Fawcett and F. Provost, "Adaptive Fraud Detection," *Data Mining and Knowledge Discovery*, vol. 1, pp. 291-316, 1997/09/01 1997.
- [13] R. Akbani, S. Kwek, and N. Japkowicz, "Applying Support Vector Machines to Imbalanced Datasets," in *Machine Learning: ECML 2004*. vol. 3201, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 39-50.

- [14] W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen, "Effective detection of sophisticated online banking fraud on extremely imbalanced data," *World Wide Web*, vol. 16, pp. 449-475, 2013/07/01 2013.
- [15] C. D'Elia, C. Marrocco, M. Molinara, G. Poggi, G. Scarpa, and F. Tortorella, "Detection of microcalcifications clusters in mammograms through TS-MRF segmentation and SVM-based classification," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 742-745 Vol.3.
- [16] M. R. Chandraratne, "Comparison of Three Statistical Texture Measures for Lamb Grading," in *Industrial and Information Systems, First International Conference on*, 2006, pp. 513-518.
- [17] T. Hill and P. Lewicki, *Statistics: Methods and Applications : a Comprehensive Reference for Science, Industry, and Data Mining*: StatSoft, 2006.
- [18] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar. *Image Texture Feature Extraction UsingGLCM Approach - published at: "International Journal of Scientific and Research Publications (IJSRP)"*, Volume 3, Issue 5, May 2013 Edition.

APPENDICES

APPENDIX A

CODING FOR PRE-PROCESSING AND SEGMENTATION

```
%%Pre-processing%%
%Choose original image
[file_name file_path] = uigetfile ('.jpg');
A = imread ([file_path,file_name]);
gray=rgb2gray(A);

%Convert to binary Image
threshold = graythresh(gray);
B = im2bw(gray,threshold);
B = bwareaopen(B,30); % remove all object containing fewer than 30
pixels
se = strel('disk',2); % fill a gap in the cells
B = imclose(B,se);
C=imfill(B,'holes'); %Fill the holes
imwrite(C,'Binary Image.jpg');

%Label the connected components
[Label,Total]=bwlabel(C,8);
subplot (1,3,1),imshow(A); title('Original Image');
subplot (1,3,2),imshow(C); title('Labelled Image');
s = regionprops(Label, 'Centroid');
subplot (1,3,3), imshow(C);
hold on
    for k = 1:numel(s)
        c = s(k).Centroid;
        text(c(1), c(2), sprintf('%d', k), ...
            'HorizontalAlignment', 'center', ...
            'VerticalAlignment', 'middle');
    end
hold off

cc = bwconncomp(C,4);
L = labelmatrix(cc);

for i=1:Total;
M = (L==i);
%figure; imshow(M)
cell = im2double(gray).*M;
%figure; imshow(cell)
%Name=strcat('Object Number:',num2str(i));
imwrite(cell,['cell',num2str(i),'.jpg']);
end
```

APPENDIX B

CODING FOR FEATURE EXTRACTION USING GLCM

```

%% GLCM Feature Extraction
function [out] = prominence(glcmin,pairs)

% If 'pairs' not entered: set pairs to 0
if ((nargin > 2) || (nargin == 0))
    error('Too many or too few input arguments. Enter GLCM and
pairs.');
```

```

elseif ( nargin == 2 )
    if ((size(glcmin,1) <= 1) || (size(glcmin,2) <= 1))
        error('The GLCM should be a 2-D or 3-D matrix.');
```

```

    elseif ( size(glcmin,1) ~= size(glcmin,2) )
        error('Each GLCM should be square with NumLevels rows and
NumLevels cols');
```

```

    end
elseif (nargin == 1) % only GLCM is entered
    pairs = 0; % default is numbers and input 1 for percentage
    if ((size(glcmin,1) <= 1) || (size(glcmin,2) <= 1))
        error('The GLCM should be a 2-D or 3-D matrix.');
```

```

    elseif ( size(glcmin,1) ~= size(glcmin,2) )
        error('Each GLCM should be square with NumLevels rows and
NumLevels cols');
```

```

    end
end

format long e
if (pairs == 1)
    newn = 1;
    for nglcm = 1:2:size(glcmin,3)
        glcm(:, :, newn) = glcmin(:, :, nglcm) + glcmin(:, :, nglcm+1);
        newn = newn + 1;
    end
elseif (pairs == 0)
    glcm = glcmin;
end

size_glcm_1 = size(glcm,1);
size_glcm_2 = size(glcm,2);
size_glcm_3 = size(glcm,3);

% checked
out.cprom = zeros(1,size_glcm_3); % Cluster Prominence: [2]

glcm_sum = zeros(size_glcm_3,1);
glcm_mean = zeros(size_glcm_3,1);
glcm_var = zeros(size_glcm_3,1);

u_x = zeros(size_glcm_3,1);
u_y = zeros(size_glcm_3,1);
s_x = zeros(size_glcm_3,1);
s_y = zeros(size_glcm_3,1);

```

```

% checked p_x p_y p_xplusy p_xminusy
p_x = zeros(size_glcm_1,size_glcm_3); % Ng x #glcms[1]
p_y = zeros(size_glcm_2,size_glcm_3); % Ng x #glcms[1]
p_xplusy = zeros((size_glcm_1*2 - 1),size_glcm_3); %[1]
p_xminusy = zeros((size_glcm_1),size_glcm_3); %[1]

% checked hxy hxy1 hxy2 hx hy
hxy = zeros(size_glcm_3,1);
hxy1 = zeros(size_glcm_3,1);
hx = zeros(size_glcm_3,1);
hy = zeros(size_glcm_3,1);
hxy2 = zeros(size_glcm_3,1);
corm = zeros(size_glcm_3,1);
corp = zeros(size_glcm_3,1);

for k = 1:size_glcm_3
    glcm_sum(k) = sum(sum(glcm(:,:,k)));
    glcm(:,:,k) = glcm(:,:,k)./glcm_sum(k); % Normalize each glcm
    glcm_mean(k) = mean2(glcm(:,:,k)); % compute mean after norm
    glcm_var(k) = (std2(glcm(:,:,k)))^2;

    for i = 1:size_glcm_1
        for j = 1:size_glcm_2
            p_x(i,k) = p_x(i,k) + glcm(i,j,k);
            p_y(i,k) = p_y(i,k) + glcm(j,i,k);
            p_xplusy((i+j)-1,k) = p_xplusy((i+j)-1,k) +
glcm(i,j,k);
            p_xminusy((abs(i-j))+1,k) = p_xminusy((abs(i-j))+1,k)
+...
                                glcm(i,j,k);
        end
    end
end

i_matrix = repmat([1:size_glcm_1]',1,size_glcm_2);
j_matrix = repmat([1:size_glcm_2],size_glcm_1,1);
i_index = j_matrix(:);
j_index = i_matrix(:);
xplusy_index = [1:(2*(size_glcm_1)-1)]';
xminusy_index = [0:(size_glcm_1-1)]';
mul_contr = abs(i_matrix - j_matrix).^2;
mul_dissi = abs(i_matrix - j_matrix);

for k = 1:size_glcm_3 % number glcms

    u_x(k) = sum(sum(i_matrix.*glcm(:,:,k)));
    u_y(k) = sum(sum(j_matrix.*glcm(:,:,k)));

    s_x(k) = (sum(sum( ((i_matrix - u_x(k)).^2).*glcm(:,:,k)
)))^0.5;
    s_y(k) = (sum(sum( ((j_matrix - u_y(k)).^2).*glcm(:,:,k)
)))^0.5;

    out.cprom(k) = sum(sum(((i_matrix + j_matrix - u_x(k) -
u_y(k)).^4).*... glcm(:,:,k)));

    glcmk = glcm(:,:,k)';
    glcmkv = glcmk(:);

```

```
hxy1(k) = - sum(glcmkv.*log(p_x(i_index,k).*p_y(j_index,k) +  
eps));  
hxy2(k) = - sum(p_x(i_index,k).*p_y(j_index,k).*...  
log(p_x(i_index,k).*p_y(j_index,k) + eps));  
hx(k) = - sum(p_x(:,k).*log(p_x(:,k) + eps));  
hy(k) = - sum(p_y(:,k).*log(p_y(:,k) + eps));
```

```
end
```


APPENDIX D

CODING FOR GUI MENU

```
%%GUI menu for Mitotic Cell Recognition%%
clear all;
close all;
clc;

while (1==1)
    UIControl_FontSize_bak = get(0, 'DefaultUIControlFontSize');
    set(0, 'DefaultUIControlFontSize', 12);
    choice=menu('Testing for mitotic cell',...
               'Test a single cell',...
               'Test a whole image',...
               'Exit');
    set(0, 'DefaultUIControlFontSize', UIControl_FontSize_bak);
    if (choice == 1)
        cell = test1cell;
    end
    if (choice == 2)
        A = TestAll;
    end
    if (choice == 3)
        clear all;
        clc;
        close all;
        return;
    end
end
end
```



```

        'HorizontalAlignment', 'center', ...
        'VerticalAlignment', 'middle');
    end
hold off

cc = bwconncomp(C,4);
L = labelmatrix(cc);

for i=1:Total;
M = (L==i);
%figure; imshow(M)
cell = im2double(gray).*M;
%figure; imshow(cell)
%Name=strcat('Object Number:',num2str(i));
imwrite(cell,['cell',num2str(i),'.jpg']);
end

%Call the Images
NumberOfImage=Total; %chose the number of images you want to give
input
prefix_image='cell'; %change the desired input image name here only
fileformat='.jpg';%change the desired input image format here only

    for N=1:NumberOfImage;
        cell= imread(strcat(prefix_image,num2str(N),fileformat));

%%%GLCM function
GLCM2 = graycomatrix (cell, 'Offset', [0 1;-1 1;-1 0;-1 -1]);
stats = prominence (GLCM2,0);
t = struct2array(stats);

%%%Testing
Class = svmclassify(ImageTrain, mean(t));

if Class == 1;
    [row, col] = find(Label==N);
    %To find Bounding Box
    sx=min(col)-0.5;
    sy=min(row)-0.5;
    breadth=max(col)-min(col)+1;
    len=max(row)-min(row)+1;
    BBox=[sx sy breadth len];
    display(BBox);
    figure,imshow(A);
    hold on;
    x=zeros([1 5]);
    y=zeros([1 5]);
    x(:)=BBox(1);
    y(:)=BBox(2);
    x(2:3)=BBox(1)+BBox(3);
    y(3:4)=BBox(2)+BBox(4);
    plot(x, y, 'r-', 'LineWidth', 2);
end
end
end

```

