MITOTIC HEp-2 CELL RECOGNITION USING LOCAL BINARY PATTERN (LBP) AND k-NEAREST NEIGHBOUR (k-NN) CLASSIFIER

By

NOOR FATIHAH BINTI MAHMOD

FINAL PROJECT REPORT

Submitted to the Department of Electrical & Electronics Engineering
In Partial Fulfilment of the Requirements
For the Degree
Bachelor of Engineering (Hons) (Electrical & Electronics Engineering)

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

MITOTIC HEp-2 CELL RECOGNITION USING LOCAL BINARY PATTERN (LBP) AND k-NEAREST NEIGHBOUR (k-NN) CLASSIFIER

By

Noor Fatihah Mahmod
13287

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

_____

Noor Fatihah Mahmod

ABSTRACT

Immune system produces autoantibody either randomly or as a result of an unknown material in the body itself. A failure of the body's own defences against diseases has result to the auto immune diseases. Auto immune diseases will start attacking its own cell as they unable to differentiate between the foreign material and its own cell. An antinuclear antibody (ANA) is required as an indicator of the autoimmune process. IIF is the recommended technique for ANA detection in patient serum. IIF slides are observed by specialists reporting for the fluorescence intensity, staining pattern and looking for the mitotic cells. Indeed, the presence of the mitotic cells significantly important due to several key factors, first to guarantee the correctness of the slide preparation process and reported the staining pattern. Therefore the ability to detect mitotic cells is acquired to develop Computer-Aided Diagnosis (CAD) system in IIF to support the specialists form image acquisition up to image classification. This work aims to highlight the features of mitotic cells and to develop recognition algorithm for mitotic cell by incorporated Local Binary Pattern (LBP) and k-Nearest Neighbour to classified unlabelled image. A completed modelling of the LBP operator is proposed which is represented by its centre pixel and a local sign-magnitude transform (LDSMT). k-NN classifies unlabelled images based on the utmost majority samples in the feature space. This work involves five stages; image acquisition, pre-processing, feature extraction, distance measurement and classification.

ACKNOWLEDGEMENT

The author believes that the success of this project would not be possible without the involvement and contributions of certain parties, organizations and individuals. Therefore, the author would like to take this opportunity to express his deepest gratitude to them. First and foremost, the author would like to thank his supervisor, Ms. Zazilah Bt. May for the continuous assistance, supervision and advice that she has given throughout the entire Final Year Project implementation. She has constantly provided guidance and valuable knowledge for every stage since the planning phase of the project until this dissertation has been completed.

Special thanks as well to the co-supervisor, Dr. Josefina Barnachea Janier for her patient guidance in the technical and non-technical knowledge. The author would also like to thank Dr. Nazreen Bt Badruddin, the FYP coordinator, for conducting several seminars and briefings to assist the FYP students as well as helping to arrange the schedule for ELECTREX and Viva for students by inviting external examiners to evaluate the students. Besides that, the author wishes to express his gratitude to his colleagues and beloved family for their support and encouragement throughout the duration of the project.

Finally, the author would like to thank all parties involved once again for making this Final Year Project a success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ANA	Antinuclear Autoantibody
CAD	Computer-Aided Diagnosis
CLBP	Completed Local Binary Pattern
CLBP_C	Completed Local Binary Pattern-Centre
CLBP_S	Completed Local Binary Pattern-Sign
CLBP_M	Completed Local Binary Pattern-Magnitude
CTDs	Connective tissue diseases
EBGM	Elastic Bunch Graph Matching
ELISA	Enzyme-linked immunosorbent assay
ED	Euclidean Distance
HEp-2	Human epithelial-like cell line
IIF	Indirect Immunofluorescence
k-NN	k-Nearest Neighbour
LBP	Local Binary Pattern
LDSMT	Local sign-magnitude transforms
LDA	Linear Discriminant Analysis
MAHO	Mobile Assited Handoff
MCSVM	Multi Class Support Vector Machine
PCA	Principle Computational Analysis
RGB	Red green blue
VAR	Variance

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Auto-antibodies is produced by the immune system either randomly or results in response to a foreign material in the body itself. Auto-immune diseases arise from a failure of the body's own defences against disease, so they will start attacking its own cell as they lose the sense of ability to differentiate between the foreign material and its own cell. Indirect Immunofluorescence (IIF) with human epithelial-like cell line (HEp-2) cells has become a dominant technique for the detection of antinuclear autoantibodies (ANA) in patient serum. ANA normally refers to the material which attached to contents of the cell nucleus. The specialist or physician at the fluorescence microscope able to detect a certain fluorescence pattern on HEp-2 reporting for the autoimmune diseases such as connective tissue diseases (CTDs), primary biliary cirrhosis, and autoimmune rheumatic diseases. Along with the modern trends, automation in IIF test have evolved to date is the "preparation of slides with robotic devices performing dilution, dispensation and washing operations" [1-2].

IIF have several limitations and disadvantages since this kind of imaging modality is a subjective method. The first disadvantage of IIF is that, it results in the variability of classification as it fully depends on the knowledge and experience of both the physicians and specialist who will evaluate the sample cells. Secondly, it involves a lot of manual work [3]. Mitotic cells is introduced to the slide cells as they work as a marker to check for the accuracy of slide preparation process and provide information on the staining pattern. In [4], the presence of mitotic cells being considered as the crucial factor to detect tumour prognosis. Therefore, mitotic cells detection will be able to build up the development of Computer –Aided Diagnosis (CAD) system in IIF. The relevance of this issue is due to the capabilities of the mitotic cells. First, they can offer a full confirmation regarding the preparation of the slide whether it is true or not and since mitotic cells can be associated with several staining patterns and all of staining antigens [5] this will help to assist the doctors in clarifying the staining patterns on the decisions stage later.

This work will present the detection of mitotic cells in the image using one of the Local Binary Pattern (LBP) variants which is known as Completed Local Binary Pattern (CLBP) as the feature extraction and classified them using k-Nearest Neighbour (KNN). Further discussion on CLBP and k-NN will be explained further in the next section.

1.2 Problem Statement

Increase in the incidence of autoimmune diseases over a last few years, fairly attributable to enhanced of the diagnostic capabilities and emergent awareness of this clinical problem in general medicine have warned the clinicians to accurately perform the laboratory determinations in IIF. However, this imaging modality technique suffers from several disadvantages limiting its performances such as the readings in IIF are subjected to inter-observer variability and it involves a lot of manual work. Mitotic cells detection plays crucial role for development of CAD system however this task is performed manually which lead to subjective interpretation and thus affect the reliability of IIF diagnosis. Hence there is need for automated cell recognition.

1.3 Objectives

The following are the three objectives of this work:

1) To identify the features of mitotic cells.
2) To develop recognition algorithm for mitotic cell using Local Binary Pattern (LBP) as the feature extraction and K-Nearest Neighbour (KNN) as a classifier.
3) To assess the performance of classification algorithm and to validate accuracy of the classification results.

1.4 Scope of Study

This project will evolve on identifying the features of mitotic cells first. There are two parts for the features of the mitotic cells which are positive HEp–2 staining patterns and the representations of the positive and negative mitosis. In this work, four classes of positive HEp – 2 staining patterns will be observe namely homogenous, nucleolar, speckled and centromere. Mitotic cell may exhibit either as negative or positive mitosis which depends on how mitotic cells of HEp–2 substrate exhibit the fluorescence patterns. In line with the second objective to develop the

recognition algorithm for mitotic cell, CLBP will be implemented to extract the feature of the mitotic cells. A completed modelling of the LBP operator is proposed which is represented by its center pixel and a local sign-magnitude transform (LDSMT). LDSMT decomposes the image into the sign and magnitude components which are complementary to each other. Next, k-NN classifier is chosen as the classifier due to its good classification performance. k-NN classifies unlabelled images based on the highest majority samples in the feature space.

1.5 Relevancy of the Project

This project is very relevance and feasible to be carried out as a Final Year Project (FYP), as mostly at the initial stage it involves with the gathering all the useful information and then to carried out the simulation part by part in order to achieve several target or output of the project. This project is very much relevant to my 4 years of undergraduate study majoring in electrical and electronics engineering. It is mainly dealing with image processing which is included in the core syllabus under data signal processing. Besides technical knowledge in programming, project management skill like time management and interpersonal communication skills are required. Also, this project challenges critical, analytical, innovative and creative thinking, which are all highly demanded in real working environment of a professional, competent and qualified engineer.

1.6 Feasibility of the Project

Feasibility of project can be estimated via two aspects, which are technical and time economic based. Technical feasibility is defined as the workability by considering the technical aspects in learning, assimilating and applying the knowledge learnt. For example, extensive literature review is conducted to look for the most realistic approaches and techniques used for the feature extraction using variants of LBP before attempting them in implementation on MATLAB. Besides that, the platform of coding development is MATLAB which is not common among engineering student. Basically, this project is feasible technically and it is simulation based. The software is downloaded online. Hence, this project is feasible technically and time and economic based.

CHAPTER 2

LITERATURE REVIEW

2.1 Medical Context

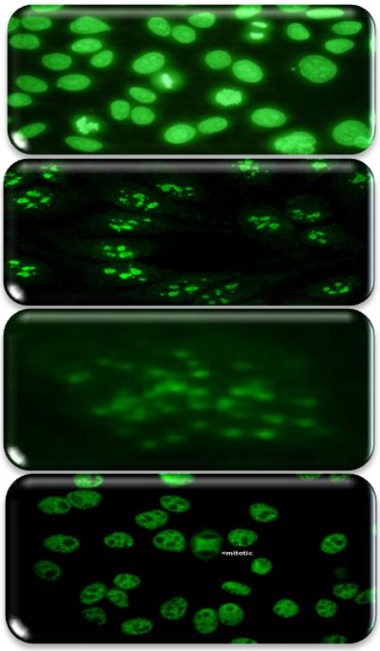
Antinuclear autoantibodies (ANAs) play an important role in the diagnosis of autoimmune diseases. ANAs are steered against a variety of nuclear antigens and can be detected in patient serum through laboratory test. The normal tests utilized for discovering and quantifying ANAs are indirect immunofluorescence (IIF) and enzyme-linked immunosorbent assay (ELISA). IIF is considered as the primary standard for ANA testing.

The most recommended method in testing ANA is IIF which was first used in 1957 but it suffered from several disadvantages limiting its efficiency and robustness. ELISA or other techniques can be used for the detection of autoantibodies entities but IIF patterns permit a general decision on the presence of the autoantibodies [5]. IIF is a subjective method which relying on the field knowledge of the specialists [6]. Pathologists, who will examined the slide cells, addressing the outcomes such as fluorescence intensity, staining pattern and reporting for the mitotic cells.

HEp-2 cell contain larger nucleus and most importantly plenty of mitotic cells that will assist during the decision stage for interpretation of the ANA pattern. Detection of the mitotic cells eventually acts as a platform stage in motivating the development of the CAD system. Mitotic cells can be modelled as the confirmation of the correctness of the tests if able to detect the presence of the mitotic cells in slide preparation process and thus assist the doctors in decision making process on the staining pattern [7].

It is worth observing the fluorescence intensity influences the execution of the consequent stages (mitotic cell recognition and staining pattern classification) since the intermediate intensity examines generally show low contrast, making the classification task harder indeed, for specialists. The most clinically practical ANA pattern is shown in the Table 1 below.

Table 1 : Example of Positive Staining Patterns [5-8]



- **Homogenous** - At the interphase cell, diffuse staining of the entire nucleus.
- **Nucleolar** - Staining only of the nucleoli
- **Centromere** - Several discrete speckled (`40-60) distributed evenly in the nucleus.
- **Speckled** - A fine or coarse granular nuclear staining in nucleus

Table 1 show positive HEp–2 samples may represent various patterns of fluorescent staining that are applicable to diagnostic purposes. Examples of each fluorescence patterns of the associated group are shown in the Table 1.

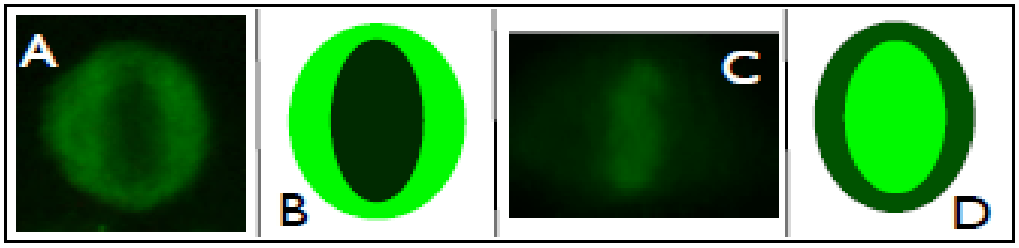


Fig.1 : Representations of Negative (Panels A-B) and Positive (Panels C-D) mitosis [7]

Mitotic cell may exhibit as positive or negative mitosis which depends on how mitotic cells of HEp–2 substrates exhibit two fluorescent patterns. Panel A and B in Fig.1 shows that the mitotic cell exhibit negative mitosis as the cell body is fluorescent while the collapse chromosome mass of the cell does not exhibit fluorescence pattern. Positive mitosis is the vice versa of the negative mitosis as shown in the Panel C and D where the collapsed chromosome mass exhibit the fluorescence pattern while cell body is weakly or does not exhibit the fluorescence pattern [7].

2.2 Dataset Descriptions

In this part, provide some information regarding on the dataset used in this work. The MIVIA HEp-2 images dataset is an outcome of the research project "Classification of Immunofluorescence Images for the Diagnosis of Autoimmune Diseases", which is available for any research purposes and can be obtained from the web page in [9]. The dataset contain the images from slides of HEp-2 substrate at the fixed dilution of 1:80 which is recommended by the guidelines. HEp-2 images were acquired using a florescence microscope (40-fold magnification) coupled with a 50W mercury vapour lamp and with a digital camera. The camera has a CCD with squared pixel of equal side to 6.45 μm. The images have a resolution of 1388×1038 pixels, a colour depth of 24 bits and they are stored in an uncompressed format.

Each cell of the dataset is manually segmented and annotated by the specialist, provided information on the staining pattern, fluorescence intensity and mitosis phase. For each segmented image, some information is included such as the staining pattern, mitotic cell, the patterns whether it is positive or negative, intensity, and coordinates of the seed object. The dataset whose composition is summarize in Appendix I, contains 28 images which almost equally distributed with respect to different patterns (6 images belonging to the centromere, 5 homogenous and coarse speckled images and 4 images belonging to the cytoplasmic, fine speckled and nucleolar). The number of cells in each image is reported in the fourth column of the table. In this work, for experimental purposes, the dataset will be divided into two categories, training set and testing set.

2.3 Feature Extraction and Selection

In this section explained in detail the selection of Local Binary Pattern (LBP) as the powerful texture operator to describe the local image. One of the main focuses of this work is the textural features as LBP is a powerful operator for texture analysis. Textural features able to distinguish various kinds of patterns [8]. In order to identify a various kind of textural features, LBP operator is introduced in this paper. LBP is a famous texture measure originally proposed by Ojala [10].

Local Binary Pattern (LBP) was introduced on 1996, and since then a lot of research have been conducted parallel to its significant value in image retrieval, object recognition and texture recognition. LBP is a well-known robust operator, due

to its important properties that are invariant to any monotonic changes in gray-scale and apply a simple computational method [11-13]. Other than that, LBP operator is also addressing as a unifying approach to texture analysis due to its powerful measure in texture analysis.

2.3.1 LBP and LBP Variants

LBP is proved to provide the best retrieval performance and also has out-perform others method such as Principle Computational Analysis (PCA), Linear Discriminant Analysis (LDA), and Elastic Bunch Graph Matching (EBGM). Due to all of these advantages, LBP has become a central attraction in recent years. Due to this reason, a lot of LBP variants are introduced. In [14], shows that a comprehensive set of 16 LBP variants with their modifications leading to a total of 38 different texture descriptors.

I. (Basic) Local Binary Patterns (LBP) [14-15]

The basic LBP was first introduced by Ojala et al. in 1996. This basic LBP works with eight neighbouring pixels based on 3×3 neighbourhood. This eight neighbouring pixel will produce 256 texture patterns (2^8). A neighbourhood is threshold by the gray-value of its center in order to produce a binary code that will eventually give the information on the local texture pattern. Value one (1) is assigned to the pixel when the neighbour pixel gives a higher or same value compared to the center pixel, or else it will gets a zero (0). Later, LBP was extended in a circular symmetric neighbourhood [13]. The notation (P,R) is implemented, where P stand for the every sampling points involved on the edge of the circle and R refers to the radius from the center pixel. Interpolation is needed to find the value of P that does not located exactly at the centre of a pixel.

II. Uniform Local Binary Patterns ($LBP_{P,R}^{U2}$) [14-15]

This type of LBP was known uniform since it well defines the uniformity of the patterns. It only has at most two bitwise transitions, basically from 0 to 1 or the opposite. Examples of the two bitwise transitions are 00001100 and 11100011. With this kind of operator there are 59 patterns if we used P=8 as it implies this formula [$P(P-1)+3$].Hence this type of LBP able to reduce from 256 texture patterns into only 59 texture patterns as it only detect the important local textures. The accuracy of this uniform LBP also may exceed 90%.

III. Rotation Invariant Uniform LBP ($LBP_{P,R}^{riu}$) [15]

This type of LBP is an extension of the original LBP. The ($LBP_{P,R}^{riu}$) operator is a powerful tool measurement of the spatial (pattern) structure and ignores others property such as contrast. Spatial structure here means that it is affected by rotation but not for the gray-scale, whereas for contrast is vice versa. ($LBP_{P,R}^{riu}$), P is represent for the quantization of the angular space whereas R will addressing the spatial of the operator [15]. Basically LBP^{riu} is better compared to the LBP^{ri} as they can provide a good discrimination. By combining ($LBP_{P,R}^{riu}$) with the rotation invariant variance measure $VAR_{P,R}$ will help in optimize the performances of the ($LBP_{P,R}^{riu}$).

IV. Completed Local Binary Pattern (CLBP) [16]

Basically, CLBP can represent more discriminant information compared to the original LBP. CLBP is developed to provide robust performance for texture classification. In this work, CLBP is selected as it provides much better texture classification accuracy than other the state of the arts LBP algorithm.

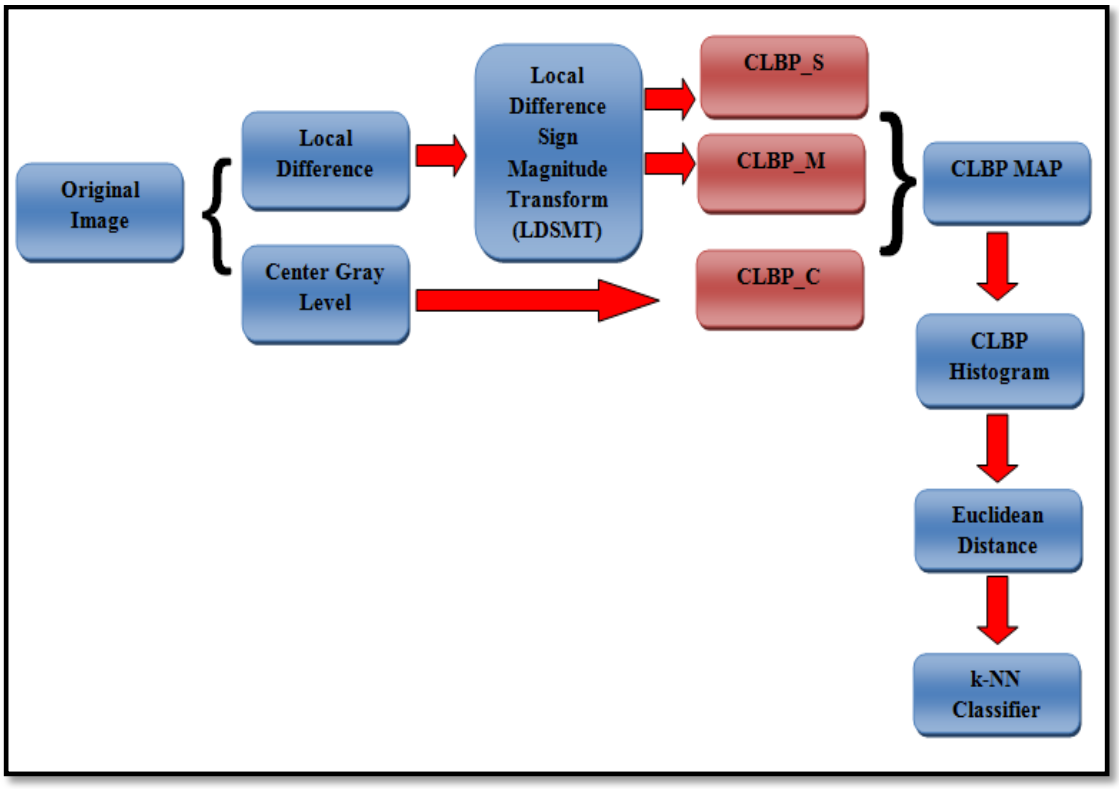


Fig.2 : Framework of CLBP

Figure 2 illustrates the whole framework of CLBP to explore all the three types of features (CLBP_C, CLBP_S, CLBP_M). The original image is represented as its center gray level (CLBP_C) and LDSMT which decompose into sign and magnitude components. Then all these three features combined to form the CLBP features map of the original image. By considering all these three features helps to effectively increase the accuracy for texture classification. Later CLBP histogram is computed and Euclidean Distance Metric is introduced to improve k-NN classifier achieve higher accuracy in texture classification.

Table 2 : Classification Rate on TC10 and TC12 Using Different Scheme [16]

	$(P,R)=(8,1)$				$(P,R)=(16,2)$				$(P,R)=(24,3)$			
	TC10	TC12		Aver -age	TC10	TC12		Aver -age	TC10	TC12		Aver -age
		"t"	"h"			"t"	"h"			"t"	"h"	
LTP	76.06	62.56	63.42	67.34	96.11	85.20	85.87	89.06	98.64	92.59	91.52	94.25
$VAR_{P,R}$	90.00	62.93	64.35	72.42	86.71	63.47	67.26	72.48	81.66	58.98	65.18	68.60
$LBP^{riu2}_{P,R}/VAR_{P,R}$	96.56	79.31	78.08	84.65	97.84	85.76	84.54	89.38	98.15	87.13	87.08	90.79
$CLBP\_S^{riu2}_{P,R}$	84.81	65.46	63.68	71.31	89.40	82.26	75.20	82.28	95.07	85.04	80.78	86.96
$CLBP\_M^{riu2}_{P,R}$	81.74	59.30	62.77	67.93	93.67	73.79	72.40	79.95	95.52	81.18	78.65	85.11
$CLBP\_M^{riu2}_{P,R}/C$	90.36	72.38	76.66	79.80	97.44	86.94	90.97	91.78	98.02	90.74	90.69	93.15
$CLBP\_S^{riu2}_{P,R}\_M^{riu2}_{P,R}/C$	94.53	81.87	82.52	86.30	98.02	90.99	91.08	93.36	98.33	94.05	92.40	94.92
$CLBP\_S^{riu2}_{P,R}/M^{riu2}_{P,R}$	94.66	82.75	83.14	86.85	97.89	90.55	91.11	93.18	99.32	93.58	93.35	95.41
$CLBP\_S^{riu2}_{P,R}/M^{riu2}_{P,R}/C$	96.56	90.30	92.29	93.05	98.72	93.54	93.91	95.39	98.93	95.32	94.53	96.26
VZ_MR8	93.59 (TC10), 92.55 (TC12, "t"), 92.82 (TC12, "h") (Average 92.99)											
VZ_Joint	92.00 (TC10), 91.41 (TC12, "t"), 92.06 (TC12, "h") (Average 91.82)											

Several studies have been conducted in order to identify which type of LBP variants will outperforms the others. Table 2 above shows the classification rate for different schemes. The result shows that all of the others schemes unable to compete with the fusion of (CLBP_C, CLBP_S, and CLBP_M). This type of scheme produces the highest accuracy in three different cases when the value of P and R differs. The range of accuracy for this experiment is at the average of 93% - 96%. Table 3 below shows the summary for the selection of feature extraction which shows that CLBP is the most powerful for the texture descriptor among LBP variants and other texture descriptors.

Table 3 : Summary of Feature Extraction Selection

PAPERS	[17]	[18]	[19]
AUTHORS	S.Singh, R.Maurya, and A.Mittal	A.Tahmasebi, H.Pourghassem and H. Mahdavi	O.Lahdenoja , M.Laiho, A.Passion
YEAR	2012	2011	2005
FEATURE EXTRACTION	Completed Local Binary Pattern (CLBP)	2D Gabor Filter	Standard Local Binary Pattern
ADVANTAGES	- One of most powerful texture descriptor among LBP variants. - Simple and efficient operator. - Outperform other feature extractor method - Low cost discriminative characteristics for facial recognition system	- Able to extract feature such as spatial locality , orientation selectivity and also some distinct features	- Standard LBP variants - Invariant with respect to monotonic gray-scale. - Outperform all the state of the art such as PCA, LDA and EBGM. - Can be extracted faster in a single scan compared to Gabor wavelets.
RESULT	- Average of the accuracy is higher compared to the other method. - Achieves accuracy of 86.4% when used with Multi Class Support Vector Machine (MCSVM) - Higher accuracy compared to the standard LBP.	- Combination of this technique with k-NN classifier produces a robust fusion technique for ear identification system. - Proposed method needed more processing time as Gabor Filter will produce a large dimension of the filtered image.	- Shows the highest accuracy for four different classes with the average accuracy of 70.8%.
APPLICATION	- Facial expression recognition - Dynamic texture recognition - Shape localization	- Ear identification system	- Industrial wood - Paper inspection - Motion detection - Face recognition

2.4 Distance Measurement

Distance measurement or dissimilarity measurement function to calculate how closes each member in the training dataset to the testing class that is being examined. In this work, the author used Euclidean Distance (ED) Metric as the dissimilarity measurement. Distance metric algorithm is introduced to improve the classification process. In this work, ED is chosen as in [20] stated that most k-NN classifiers use ED to measure the distance between two images as ED is simple and mathematically convenient metric. Table 4 below described in details the summary of distance measurement such as ED, Chi-Squared and Bhattacharya.

Table 4 : Summary of Distance Measurement

PAPERS	[16]	[21]	[22]	[23]
AUTHORS	S.Noh	M. Hafner, A. Gangl , M. Liedlgruber, A. Uhl, A. Vecsei, and F. Wrba	X.Xianchuan and Z.Qi	Z.Ruifeng, J.Hailin, and T.Zhenhui
YEAR	2012	2009	2009	2000
APPLICATION	-Shaping classification -Boundary detection	- Endoscopic imagery	- Used in medical image retrieval	- Mobile Assited Handoff (MAHO)
DISTANCE MEASUREMENT	Chi Squared distance metric	Bhattacharya Distance Metric	Euclidean Distance Metric (ED)	
ADVANTAGES	-Able to improve the accuracy of k-NN classifier. - Robust and discriminative and able to lessen the distinction caused by bins with large values.	-Always produce slightly higher classification result.	- Simple and mathematically convenient metric. - Robust to small perturbation of images. - Able to improved the performance of LBP algorithm. - can be fixed in several classification such as PCA, SVM and LDA and k-NN classifier.	-ED reasonably fit for MAHO - Decrease the redundant handoffs and handoff delay. - Consistent channel feature information to assist rate adaptation and power control.
RESULT	-Shows the least classification error for four different dataset.	- Improved the performance of the experiment.	- Improved the performance of the experiment from 89.5% to 92.5%.	- The average of metric remains consistent over diverse coded.
OTHER APPLICATION	- Face recognition, shape localization and dynamic texture recognition	- Face detection, dynamic texture recognition and shape localization	- Face recognition, ear identification system and detection of facial changes in ICU patients	

2.5 Classification Process (k-NN Classifier)

There are several types of classifiers such as support vector machine (SVM) classifier, AdaBoost classifier [24], and Random Forest Classifier. AdaBoost classifier is being categorized as a statistical method and always used for object detection (details information in [24]).One of the most good classification performance of the classifier is k-Nearest Neighbour (KNN) [25].

k-NN will save all the information of the training database then will be carried out a classification only if the testing objects match with either one in the training samples exactly. In this paper after the images of the cell is being transformed using CLBP, later histograms will be created based on the CLBP data and finally k-NN classifier will be conducted with a suitable histogram distance metric such as Bhattacharya, chi-squared and Euclidean distance (ED). In [26], Bhattacharya distance metric is being computed. In this work, ED is incorporated with the k-NN classifier.

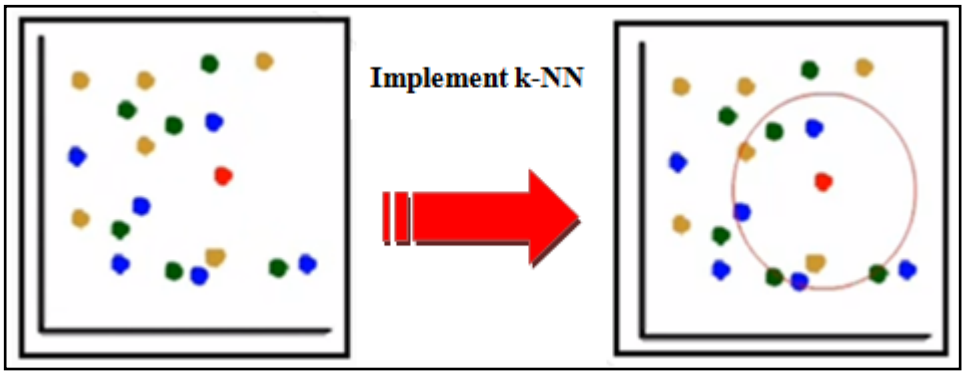


Fig.3 : k-NN Rule Classifier

Figure 3 above illustrates how k-NN classifier works. ‘k’, is the number of nearest neighbours where this parameter can be chosen to optimize the result of classification [24]. For example in order to find ‘k’ (e.g. k=6) nearest neighbours of query point (closest point to the red point), in Figure 3 shows that most of nearest or majority points (3) are from blue class, hence blue class will be assign to query structure. Basically, KNN classifies unlabelled images based on the highest majority samples in the feature space.

2.5.1 Application using k-NN Classifier

I) An Ear Identification System

In [27], present the paper work for image retrieval in ear recognition system. It used Local-Gabor as the feature extraction to improve the texture of an ear and applied (k-NN) as the classifier. In the evaluation part, to obtain the highest performance, the value of 'k', nearest neighbour need to be manipulated so that certain desired level of accuracy can be achieved. The result is tabulated, and it shows that the suitable values of (k), for this experiment are in between 12-20.

II) Detection of Facial Changes for ICU Patients Using KNN Classifier

This paper [28], explained on how to classify each facial changes of patient in Intensive Care Unit (ICU), using k-NN as the classifier. The patient in ICU need to have a proper monitoring by the nurses, however the nurses might miss to notice any undesired event in the ICU. In [23], proposed to design a system to monitoring the eyes of the patient as k-NN classifier can easily detected the movement of eye and mouth of the patient. The result shows that, k-NN classifier have achieved an accuracy with the average 94% (k=1).

2.6 Project Framework

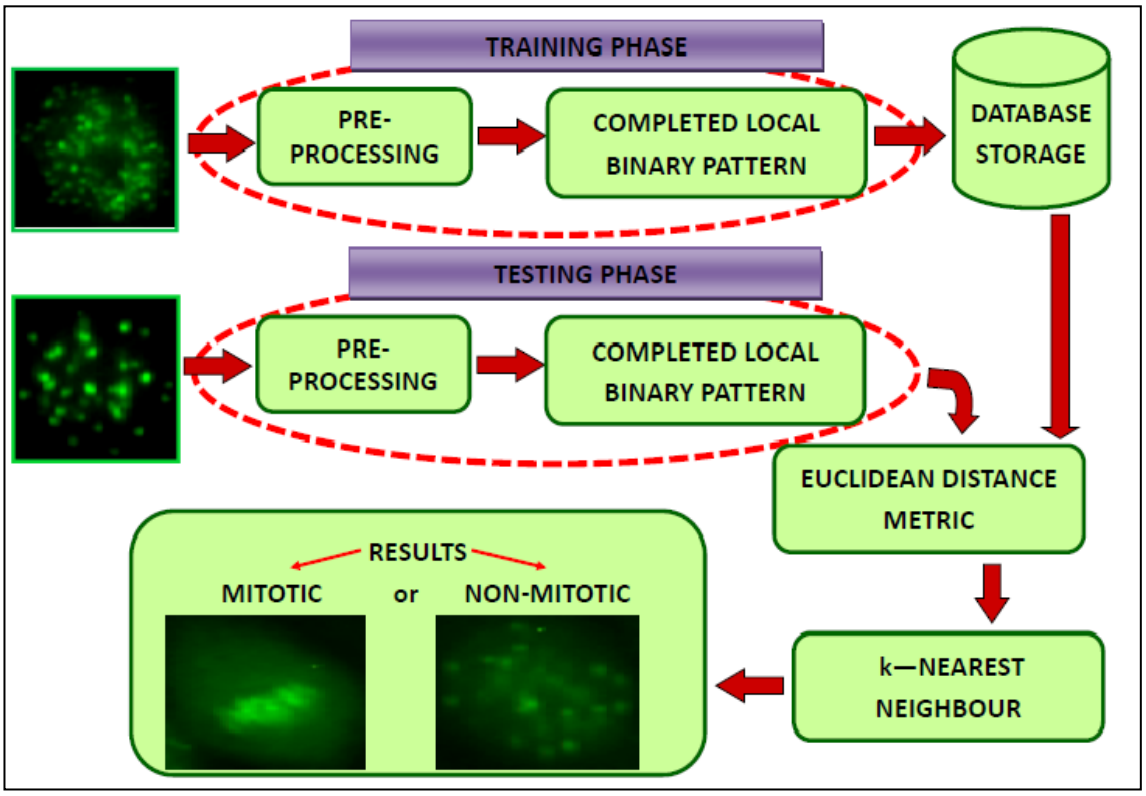


Fig.4 : Illustration of the Project Framework

Figure 4 illustrates the framework of the whole project, which will be executed in the MATLAB environment. The input image will first undergo the pre-processing part. In this part the RGB image will be transform into the gray-scale image by removing the hue and any saturation while preserves the luminance. The noise also will be reduced while retaining and maintaining the shape and contrast of the cell.

For the feature extraction, LBP operator (CLBP) will be used and for the classifier k-Nearest Neighbour (k-NN) will be implemented. Based on the framework above, there are two phases involved. Training phase and the verification phase. The k-NN classifier will retain information of all the training data in the database storage and it will performs classification once the features of the test image in the verification phase match with the one of the training data exactly stored in the database storage.

CHAPTER 3

METHODOLOGY

3.1 Image Processing Flow

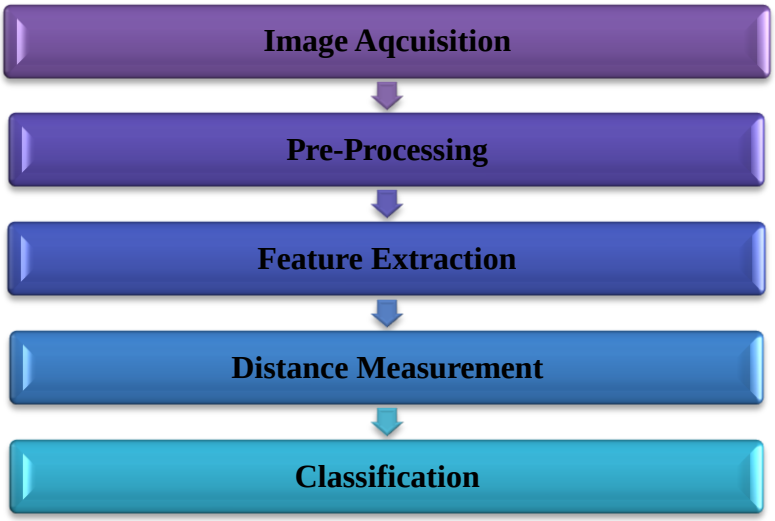


Fig.5 : Image Processing Flow

Figure 5 presents a flowchart of the proposed method in the image processing flow. The image processing flow starts with the image acquisition where the image of mitotic and non-mitotic HEp-2 cell will be use. The dataset used for this activity is publicly available mainly for research purposes. In the pre-processing part the RGB image is transformed into the gray-scale by removing the hue and saturation while preserves the luminance. Pre-processing is an important step before the feature extraction as IIF images that have noises will make the feature extraction process become arduous [29-31].

Next is the feature extraction where CLBP is introduced here in order to explore all three types of features (centre, sign and magnitude) which can provide information on the local difference. Then, distance measurement or dissimilarity measurement is incorporated in order to improve the accuracy of the classifier. In this work Euclidean Distance Metric is designed to improve the k-NN classifier. In the classification process, k-NN classifier will classify unlabelled images to its respected classes. k-NN classifier rule based on the highest majority point to the testing data.

3.2 Project Flowchart (FYP I & FYP II)

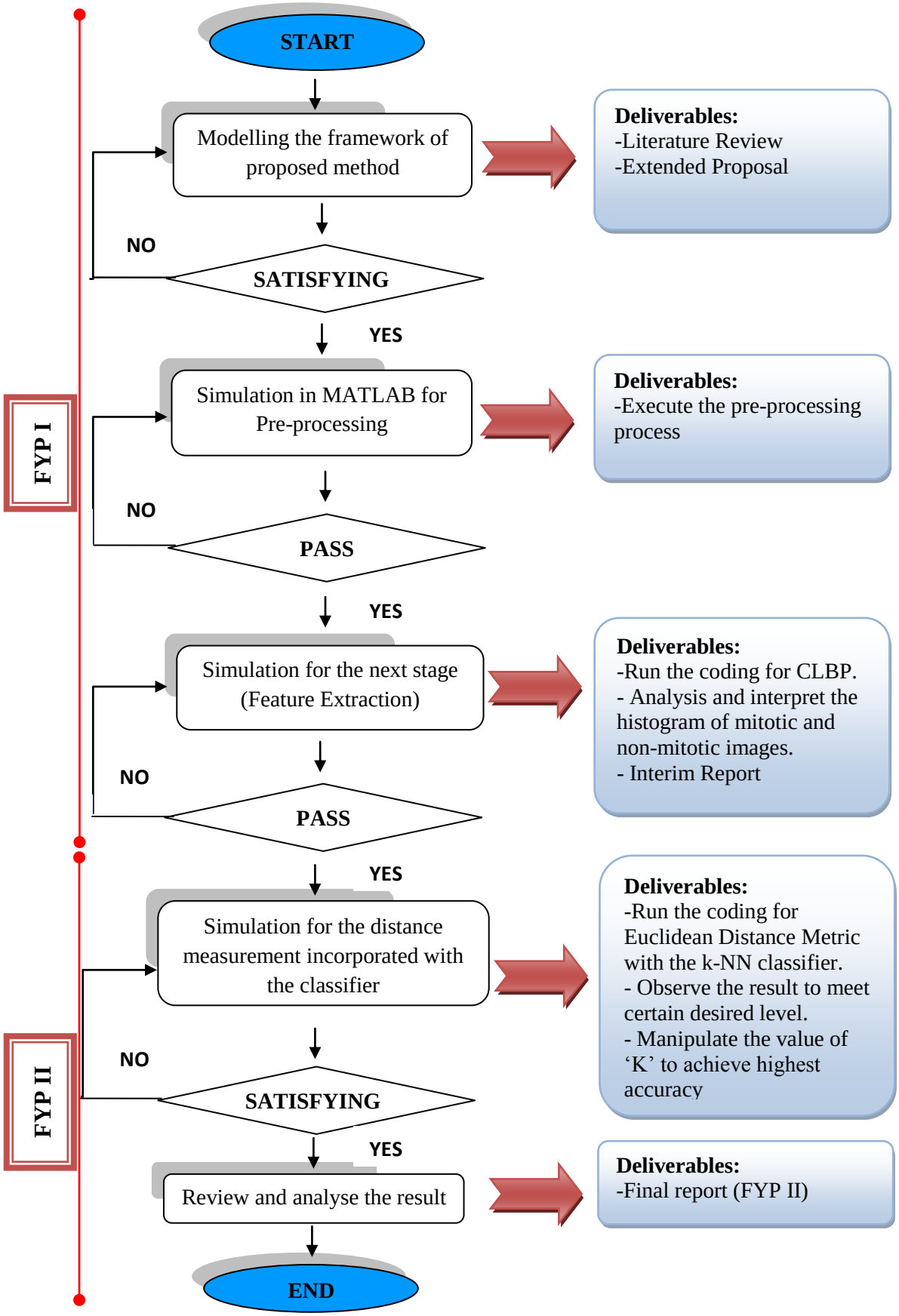


Fig. 6 : Project Flowchart

3.3 Project Activities

1. RESEARCH AND STUDY

Research on the mitotic HEp-2 cells, Local Binary Pattern (LBP) and k-Nearest Neighbour (k-NN) is conducted. Study in details for each part of it. Collect all the information regarding the topic and choose the most related one as the references

2. PROPOSAL WRITING

Clarify the problem statement and the objective of this project. Focus on the title by narrow down the scope of study in order to increase the relevancy and feasibility of this project.

3. SIMULATION

The simulation is conduct by using the MATLAB software. There are two parts of coding need to be executed in the simulation which are training phase and verification phase

4. DESIGN IMPROVEMENT

If the accuracy of the results does not meet certain desired level, any improvement should be carried out to enhance the effectiveness of the project. Result analysis will be conducted if the required output meets the desired level in the project.

5. RESULT ANALYSIS

The final result should be that the feature extraction of CLBP and the classifier k-NN able to recognise the mitotic and non-mitotic HEp-2 cells at the verification phase. Any recommendation is made for further improvement.

Fig.7 : Details of Project Activities

3.4 Gantt Chart & Key milestone

Table 5 : Timelines for FYP I (SEMESTER 1)

NO	DETAIL WORK	WEEK													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Title selection and confirmation	■	■												
2	Preparation of extended proposal			■	■	■									
3	Extended proposal submission						◆								
4	Preparation for proposal defence							■	■						
5	Proposal defence and progress evaluation									◆					
6	Preparation of interim report										■	■			
7	Draft of interim report submission												◆		
8	Improvement of interim report													■	
9	Interim report submission														◆

■ Process ◆ Suggested Milestone

Table 6: Timelines for FYP II (SEMESTER 2)

NO	DETAIL WORK	WEEK														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Simulation of feature extraction phase															
2	Simulation of k-Nearest Neighbour (k-NN) classifier															
3	Preparation of progress report															
4	Progress report submission															
5	Project work continues															
6	Pre EDX															
7	Submission of Draft Report															
8	Submission of Dissertation (Soft bound) and Technical Paper															
9	Submission of Technical Paper															
10	Oral presentation (Viva)															
11	Submission of Dissertation (Hard bound)															

Process Suggested Milestone

Tables 5 and 6 present key milestone and the Gantt chart for FYP I and FYP II. For FYP 1 all the four milestones already being completed starting with extended proposal submission, following with proposal defence presentations and towards the end draft interim and final interim report submission. In FYP II several key milestones must be achieved in order to complete the work. Table 6 shows there are total of five submissions to be made for FYP II, starting with progress report submission, following with submission of draft report, and towards the end submission of dissertation (soft bound), submission of technical paper and also submission of dissertation in hard bound.

3.5 Simulation Software

To carry on the simulation of this project the only software needed is MATLAB. The pre-processing part, segmentation, feature extraction and classification will be implemented in the MATLAB. As the focused on this paper mainly on the CLBP and k-NN classifiers, a detailed coding of the MATLAB will be introduced here.

CHAPTER 4

RESULTS & DISCUSSION

In this work, the author attempt to classify the mitotic and non-mitotic cell which exhibit one of the staining patterns. The databases consist of two separate phase which are training and testing phase. Twelve images were randomly picked as training images (6 mitotic cells and 6 non-mitotic cells) and another hundred images were randomly chosen as testing images (5 mitotic cells and 95 non-mitotic cells) which exhibit the centromere pattern. All of these images were randomly picked to test the reliability of the algorithm developed.

The coding provided in the appendixes shows four files .m from which were used in the MATLAB environment. The pre-processing part is encapsulated in a function file, TrainTest file is the one that uses all the other three files (Pre-processing, clbp, get-mapping) and train the system with (training images) N samples and at the end test it using k-NN. Below, shows some images of mitotic and non mitotic cells that were used as a training database. Figures 8 and 9 shows non-mitotic and mitotic image (centromere pattern), the right side of the cell image is the result of the image that undergone pre-processing part.

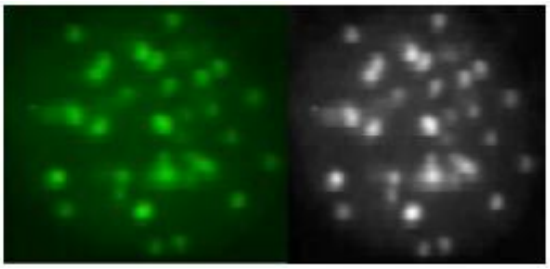


Fig.8 : Example of Non-mitotic Image

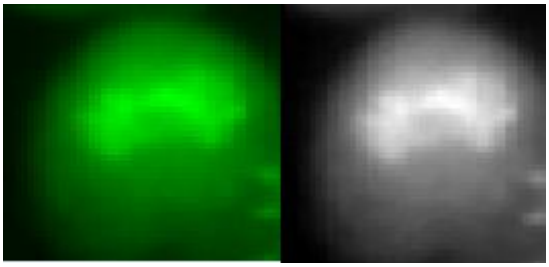


Fig.9 : Example of Mitotic Image

After pre-processing stage, CLBP will extract these images into three components, CLBP_C, CLBP_M and CLBP_S. The CLBP codes are computed using N sampling points on a circle of radius R and using mapping table defined by MAPPING. Getmapping.m will return a structure containing a mapping table for LBP codes. There are three possible values for mapping type, 'u2' for uniform LBP which consists only two number of transitions, 'ri' for rotation invariant and 'riu2' for uniform rotation-invariant LBP. In this work, as the CLBP use the rotation invariant uniform pattern, the amount of pattern matching is reduced from 256 bins to 10 bins only, based on the formula below when P = 8.

(P + 2) **(1)**

Next, Euclidean Distance (ED) Metric will measure the closeness of each member in the training dataset to the testing class that is being examined. In [22] stated that ED able to significantly improved the performance of LBP algorithm. Hence, the combination of CLBP and ED in this works help to increase the accuracy of the experiments. ED refers to the distance between two points on the real line, which is the absolute value of their numerical difference if in the one dimension. Thus if x and y is two points on the real line, then the distance between them is known by the formula below:

$$\sqrt{(x-y)}^{2} = |x\text{-}y| \qquad (2)$$

The ED first calculated between the mitotic sub-set and the testing image (for the Sign and Magnitude histogram) and then for non-mitotic sub-set and the testing image (for the Sign and Magnitude histogram). Then k-NN is applied, for example in this experiment value of 'k' is 1. The value of 'k' can be any number but must be within the number of training images (N) which means (neighbours <= N). For two classification problem, it is helpful to choose 'k' to be an odd number as this avoid tied votes. Values of 'k' need to be manipulate to get the best result. If the datsaset is not very large then another method to choose number of neighbour or 'k' is using Cross-validation. For each patterns two histograms (sign and magnitude histogram) are calculated then each histogram undergoes a k-NN classifier and make its own decision. If the image is classified, and shows that sign and magnitude histogram

display different results, (as shown in screenshot Figure 10) this means that the classifier unable to discriminate between the two classes (mitotic and non-mitotic cells).

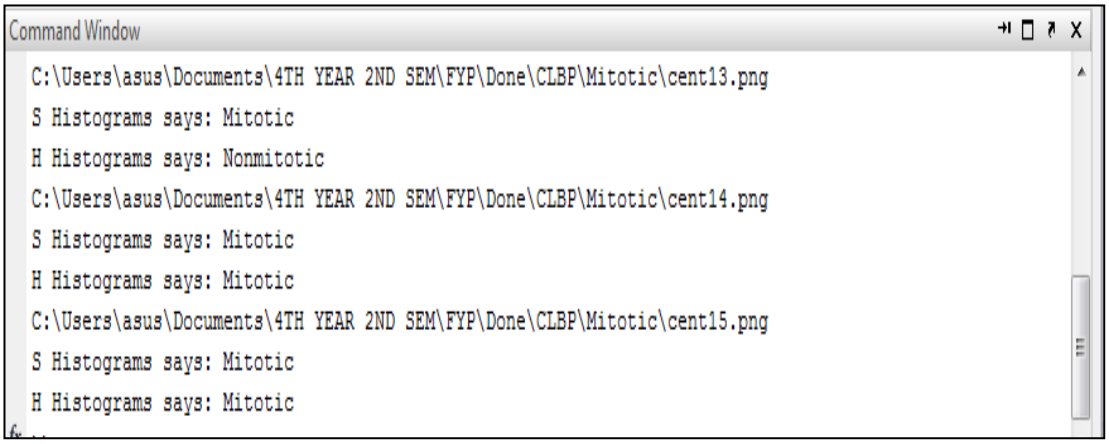


Fig.10 : Example of Classification Result for Testing Images

4.1 Accuracy Rate in Classification Process

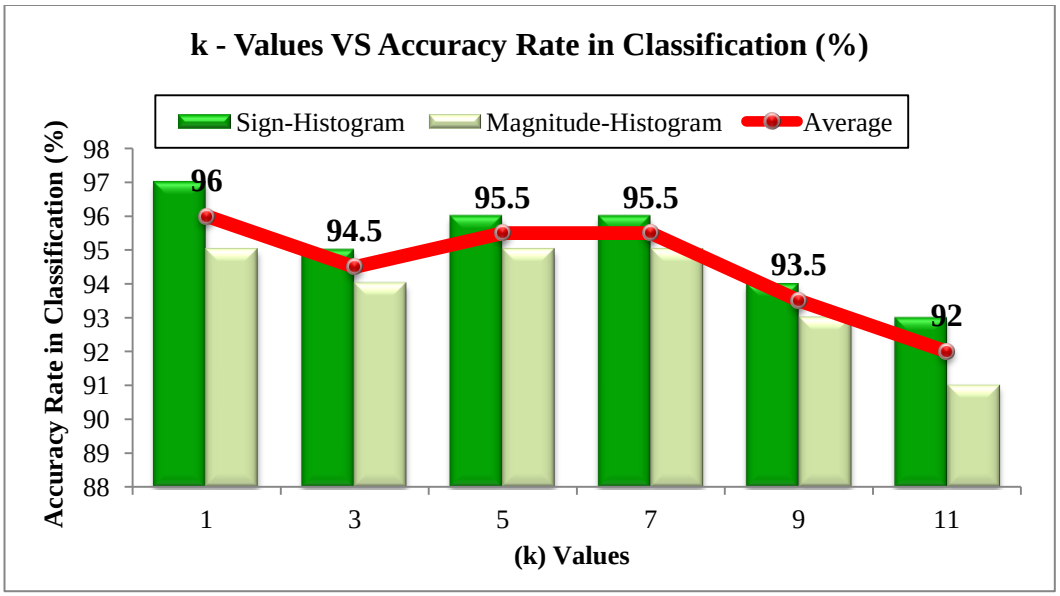


Fig.11 : k-Values VS Accuracy Rate in Classification (%)

Figure 11 above shows the result for the k-values versus accuracy rate in classification process. The value of 'k' is manipulated from 1 until 11 in order to achieve highest accuracy. Based on this graph clearly shows that sign histogram shows higher accuracy compared to the magnitude histogram as sign preserve much information on local difference compared to the magnitude component. The highest

accuracy are 97% (sign-histogram) and 95% (magnitude-histogram) when ‘k’=1. The accuracy of this experiment can further be increased by the increase in number of training images for each database.

4.2 Reliability of Algorithm Developed

Sensitivity or also known as percentage coverage of positive (mitotic cell) is the percentage of positive example predicted as positive. A sensitivity of 100% means that the test recognizes all mitotic cell as such. Sensitivity alone does not tell us how well the test predicts other classes (refer to the negative cases). In the binary classification, as illustrated below, this is the corresponding _specificity test, or equivalently the sensitivity for the other classes.

$$\textbf{Sensitivity} = \frac{TP}{TP+FN} \times 100 \qquad (3)$$

Specificity or also known as percentage coverage of negative (non-mitotic) examples predicted as negative. A 100% of specificity means that the test recognizes all non-mitotic cells as non mitotic. Specificity alone does not tell us how well the test recognizes positive cases. We also need to know the sensitivity of the test to the class or equivalently the specificities to the other classes.

$$\textbf{Specificity} = \frac{TN}{FP+TN} \times 100 \qquad (4)$$

Accuracy is the degree percentage of correctly predicted according to their respective classes (mitotic and non-mitotic cell).

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (5)$$

- TP = Mitotic cell correctly identified as mitotic cell
- FP = Non-mitotic cell wrongly identified as mitotic cell
- TN = Non-mitotic cell correctly identified as non-mitotic cell
- FN = Mitotic cell wrongly identified as non-mitotic cell

4.2.1 Sign Histogram

Table 7 : Criteria of classification of a prediction into TP, TN, FP and FN

		Actual	
Predicted		**Positive (Mitotic cell)**	**Negative (Non-mitotic cell)**
	Positive (Mitotic cell)	**TP** = 5	**FP** = 3
	Negative (Non-mitotic cell)	**FN** = 0	**TN** = 92
		Sensitivity = (5/(5+0)) × 100 = 100 %	**Specificity =** (92/(92+3)) × 100 = 96.84%

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$= ((5+92)/(5+92+3+0)) \times 100\%$$

$$= 97\%$$

4.2.2 Magnitude Histogram

Table 8 : Criteria of classification of a prediction into TP, TN, FP and FN

		Actual	
Predicted		**Positive (Mitotic cell)**	**Negative (Non-mitotic cell)**
	Positive (Mitotic cell)	**TP** = 3	**FP** = 3
	Negative (Non-mitotic cell)	**FN** = 2	**TN** = 92
		Sensitivity = (3/(3+2)) × 100 = 60 %	**Specificity =** (92/(92+3)) × 100 = 96.84%

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$= ((3+92)/(3+92+3+2)) \times 100\%$$

$$= 95\%$$

Based on the results of accuracy above clearly proved that sign histogram (97%) provide higher accuracy compared to the magnitude histogram (95%) as sign will preserve more information of local difference than magnitude. Based on the experimental results shows in Figure 11 implies that texture classification using sign features achieves much higher performance than using the magnitude features. In addition, it will also been seen that by coding both the sign and magnitude features into rotation invariant binary codes and then fusing them, resulted in much better texture classification marks rather than using only one of them [16].

4.3 10-Fold Cross Validation Technique

Cross validation technique also being implemented in this experiment due to the limited number of mitotic cell images that have made the training dataset become small. The advantages of this cross validation technique is that all dataset are eventually used for both training and testing and it is very useful for overcoming the problem of overfitting when the size of dataset is small. Overfitting occurs when a model requires more information than the data can provide. It usually occurs because the criterion used for training the model is not the same as the criterion used to judge the efficacy of the model.

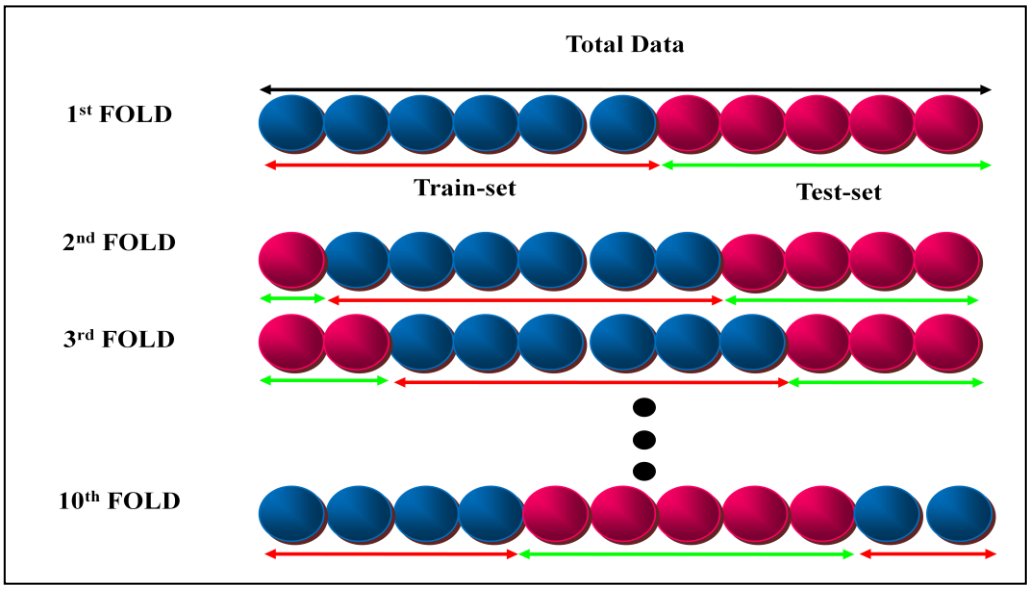


Fig.12 : 10-Fold Cross Validation Technique

In practice, the choice of the number of folds depends on the size of the dataset, for large dataset, even 3-Fold Cross Validation will be quite accurate. A common choice for K-Fold Cross Validation is 10 fold as shown in Figure 12. In this experiment for 10 fold cross validation technique, the value of training images is set

to 6 images for each class (mitotic and non-mitotic) while testing images is 5 images for each class. For this technique, 10-fold partition of dataset is created as shown in the Figure 12 above, and then the analysis is executed on each fold and average the overall error estimate.

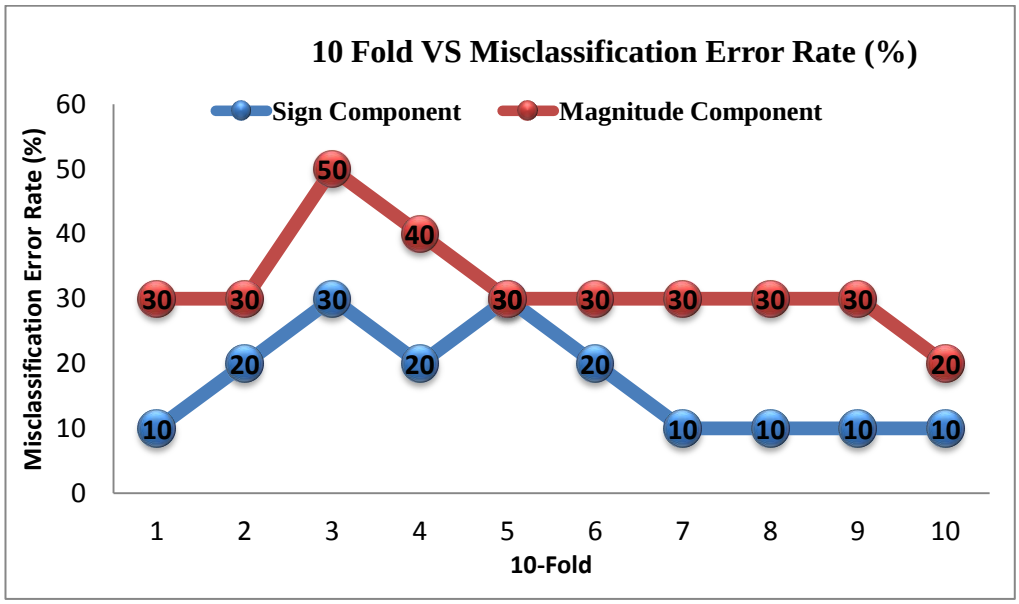


Fig.13 : 10-Fold VS Misclassification Error Rate (%)

Figure 13 above shows the line graph of 10-fold versus the misclassification error rate. Based on the result above proved that magnitude component will always shows higher error estimation compared to the sign component. The average error for sign histogram is (17%) and magnitude histogram is (32%). Based on the average error proved that if the correct data used for training dataset able to provide least error estimation such as on the first fold for sign component the total error estimate is only 10%. This also can be concluded that by enrich the dataset with a large number of training images will eventually help to increase the accuracy of the classification result as this can avoid the overfitting problem.

CHAPTER 5

CONCLUSION & RECOMMENDATION

5.1 Conclusion

As the conclusion, in line with the objectives the expected result of this project is able to identify the features of mitotic HEp-2 cells. For this part the author managed to identify the features of this cells which can be classified into four various classes according to their respective positive HEp-2 staining patterns. Mitotic cells also have different kind of representations with the non-mitotic cell.

In this paper, the author analysed LBP from a viewpoint of LDSMT which is defined by all three operator CLBP_S, CLBP_M and CLCP_C to extract the image local gray level, which are the sign and magnitude features of local difference respectively. It shows that the sign component is more important than the magnitude component in preserving the local information which explains the reason of why conventional LBP (CLBP_S) features are more effective than the CLBP_M features. In addition, it be seen that by fusing them (sign and magnitude features) much better texture classification results can be obtained rather than using only one of them.

Overall CLBP able to provide higher accuracy compared to the state-of-the-arts LBP algorithms which explained why CLBP outperformed the other feature extractor method. In this work also, the author shows how the distance metric helped to improve the accuracy of the k-NN classifier. Hence, this kind of combination able to remarkably improve the accuracy on the test set and able to outperforms the other type of classifiers.

Based on this experiment, the proposed method able to emphasize several important issues, first the key design in order to obtain high performance of classification depends on the local descriptions used to explicitly model the textures while type on classifier incorporated will only slightly improved the recognition rates [32]. Other than that, the experiments have confirmed that to classify each pattern of the image whether it is mitotic or non-mitotic are related to the level of visual

likeness to other patterns. When the patterns share an analogous visual appearance, they will often confuse with each other.

5.2 Recommendation and Future Work

In the future, the author will address several main issues. First, to observe how does the fluorescence intensity will affect the recognition rate of the classifier. The fluorescence intensity basically will significantly affect the recognition performance of the classifiers, as positive cells will have different representations with the intermediate ones as they have different features especially in term of brightness and intensity. In [32], stated that the positive cells will be recognized much more reliably compared to the intermediate ones.

Second, to analyse the impact of skew dataset to the result of the classifier. In this work the training images for both classes is similar, hence in future that has to be taken into account is the performance of the classifier towards the skew dataset during the training of the classifier. Basically, in several cases the detection systems were biased towards the larger dataset set during the training phase.

A third important issue regards the dataset. In this work, the number of images for mitotic cells is very limited which is only 11 images. Due to this problem, have limits the performance of the classifier, as the author unable to increase the number of training images for both class. In order to improve the accuracy of the classifier, a larger number for training images is required in order to avoid the problem of overfitting which is occur when model requires more information than the data can provide. Hence, in the future the author will address this issue by enrich dataset with mitotic cells to improve the classification rate between mitotic and non-mitotic cells.

CHAPTER 6

REFERENCES

[1] P. Soda, G. Iannello, and M. Vento. "A multiple experts system for classifying fluorescence intensity in antinuclear autoantibodies analysis, " in Pattern Analysis & Applications, 12(3):215–226, September 2009.

[2] A. Rigon, P. Soda, and et al. "Indirect immunofluorescence in autoimmune diseases: Assessment of digital images for diagnostic purpose, " Cytometry B (Clinical Cytometry),72:472–477, 2007.

[3] 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.

[4] C. Sommer, L. Fiaschi, F. A. Hamprecht, and D. W. Gerlich, "Learning-based mitotic cell detection in histopathological images," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 2306-2309.

[5] U. Sack, S. Knoechner, , and et al. " Computer-assisted classification of HEp-2 immunofluorescence patterns in autoimmune diagnostics, " in Autoimmunity Reviews, 2:298–304, 2003.

[6] 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*, 2009, pp. 1-4.

[7] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Early experiences in mitotic cells recognition on HEp-2 slides," in Proceedings of the 2010 IEEE 23rd International Symposium on Computer-Based Medical Systems: IEEE Computer Society.

[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*, 2012, pp. 689-694.

[9] "MIVIA HEp-2 Images Dataset, " 2010. [Online]. Available: http://mivia.unisa.it/datasets/biomedical-image-datasets/hep-2-image-dataset/

[10] T. Ojala, M. Pietik¨ainen, and D. Harwood. "A comparative study of texture measures with classification based on feature distribution", in Pattern Recognition, 29(1):51–59, 1996.

[11] S. Cuiyu, L. Peijun, and Y. Fengjie, "Multivariate Texture Measured by Local Binary Pattern for Multispectral Image Classification," in *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, 2006, pp. 2145-2148.

[12] Takala, V., Ahonen, T., Pietikäinen, M. "Block-Based Methods for Image Retrieval Using Local Binary Patterns," in: Image Analysis, SCIA 2005 Proceedings, Lecture Notes in Computer Science 3540 (2005)

[13] M. Pietikäinen, "Image Analysis with Local Binary Patterns," pp. 115-118, 2005.

[14] N. P. Doshi and G. Schaefer, "A comprehensive benchmark of local binary pattern algorithms for texture retrieval," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 2760-2763.

[15] T. Ojala, M. Pietik¨ainen, and T. M¨aenp¨a¨a. Multiresolution gray-scale and rotation invariant texture classification with local binary pattern. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, 2002.

[16] G. Zhenhua, D. Zhang, and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *Image Processing, IEEE Transactions on*, vol. 19, pp. 1657-1663, 2010.

[17] S. Singh, R. Maurya, and A. Mittal, "Application of Complete Local Binary Pattern Method for facial expression recognition," in *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*, 2012, pp. 1-4.

[18] A. Tahmasebi, H. Pourghassem, and H. Mahdavi-Nasab, "An Ear Identification System Using Local-Gabor Features and KNN Classifier," in *Machine Vision and Image Processing (MVIP), 2011 7th Iranian*, 2011, pp. 1-4.

[19] O. Lahdenoja, M. Laiho, and A. Paasio, "Local binary pattern feature vector extraction with CNN," in *Cellular Neural Networks and Their Applications, 2005 9th International Workshop on*, 2005, pp. 202-205.

[20] Blitzer, John, Kilian Q. Weinberger, and Lawrence K. Saul. "Distance metric learning for large margin nearest neighbor classification." *Advances in neural information processing systems*. 2005.

[21] M. Hafner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vecsei, and F. Wrba, "Pit pattern classification using extended Local Binary Patterns," in *Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on*, 2009, pp. 1-4.

[22] X. Xu and Q. Zhang, "Medical Image Retrieval Using Local Binary Patterns with Image Euclidean Distance," in *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, 2009, pp. 1-4.

[23] Ruifeng, Zhao, Jiang Hailin, and Tan Zhenhui. "Mobile assisted handoff based on Euclidean distance metric." Vehicular Technology Conference, 2000. IEEE VTS-Fall VTC 2000. 52nd. Vol. 4. IEEE, 2000.

[24] F. Kadlcek and O. Fucik, "Evolutionary design of local binary pattern feature shapes for object detection," in *Adaptive Hardware and Systems (AHS), 2012 NASA/ESA Conference on*, 2012, pp. 137-144.

[25] R. V. Rajesh, J. Veerappan, S. K. Sujitha, and E. A. Kumar, "Classification and retrieval of images using texture features," in *Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on*, 2012, pp. 1-5.

[26] M. Hafner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vecsei, and F. Wrba, "Pit pattern classification using extended Local Binary Patterns," in *Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on*, 2009, pp. 1-4.

[27] A. Tahmasebi, H. Pourghassem, and H. Mahdavi-Nasab, "An Ear Identification System Using Local-Gabor Features and KNN Classifier," in *Machine Vision and Image Processing (MVIP), 2011 7th Iranian*, 2011, pp. 1-4.

[28] M. Naufal Mansor, S. Yaacob, R. Nagarajan, C. Lim Sin, M. Hariharan, and M. Ezanuddin, "Detection of facial changes for ICU patients using KNN classifier," in *Intelligent and Advanced Systems (ICIAS), 2010 International Conference on*, 2010, pp. 1-5.

[29] C. Ai-Shan, L. Po-Ting, L. Yi-Ping, H. Yu-Len, and H. Tsu-Yi, "Automatic detection of antinuclear autoantibodies cells in indirect immunofluorescence images," in *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, 2010, pp. 137-140.

[30] H. Yu-Len, J. Yu-Lang, H. Tsu-Yi, and C. Chia-Wei, "Adaptive Automatic Segmentation of HEp-2 Cells in Indirect Immunofluorescence Images," in *Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC '08. IEEE International Conference on*, 2008, pp. 418-422.

[31] H. Yu-Len, C. Chia-Wei, H. Tsu-Yi, and J. Yu-Lang, "Outline Detection for the HEp-2 Cell in Indirect Immunofluorescence Images Using Watershed Segmentation," in *Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC '08. IEEE International Conference on*, 2008, pp. 423-427.

[32] Foggia, P., Percannella, G., Soda, P., & Vento, M. (2013). Benchmarking HEp-2 Cells Classification Methods. *Medical Imaging, IEEE Transactions on, PP*(99), 1-1. doi: 10.1109/TMI.2013.2268163.

APPENDICES

APPENDIX A

Composition of the HEp-2 Images Dataset

IMAGE ID	PATTERN	INTENSITY	NUMBER OF CELLS
1	Homogenous	Positive	61
2	Fine Speckled	Intermediate	48
3	Centromere	Positive	89
4	Nucleolar	Intermediate	66
5	Homogenous	Intermediate	47
6	Coarse Speckled	Positive	68
7	Centromere	Intermediate	56
8	Nucleolar	Positive	56
9	Fine Speckled	Positive	46
10	Coarse Speckled	Intermediate	33
11	Coarse Speckled	Intermediate	41
12	Coarse Speckled	Positive	49
13	Centromere	Positive	46
14	Centromere	Intermediate	63
15	Fine Speckled	Intermediate	63
16	Centromere	Positive	38
17	Coarse Speckled	Positive	19
18	Homogenous	Positive	42
19	Centromere	Intermediate	65
20	Nucleolar	Intermediate	46
21	Homogenous	Intermediate	61
22	Homogenous	Positive	119
23	Fine Speckled	Positive	51
24	Nucleolar	Positive	73
25	Cytoplasmic	Intermediate	24
26	Cytoplasmic	Positive	36
27	Cytoplasmic	Positive	38
28	Cytoplasmic	Intermediate	13

APPENDIX B

CODING

PART 1 (Pre-Processing)

```
function [SHist MHist] = Preprocessing(I)

    %% PRE-PROCESSING
% convert original image to gray image
Image = rgb2gray(I);
% display gray image in the second quadrant
subplot (1,1,1);
imshow(Image, []), title('GrayImage');

%% FEATURE EXTRACTION

mapping=getmapping(8,'riu2');
[SHist MHist] = clbp(I,1,8,mapping,'h');
end
```

PART 2 (clbp.m)

```
function [CLBP_S,CLBP_M,CLBP_C] = clbp(varargin) %
error(nargchk(1,5,nargin));

image=varargin{1};
d_image=double(image);

if nargin==1
    spoints=[-1 -1; -1 0; -1 1; 0 -1; -0 1; 1 -1; 1 0; 1 1];
    neighbors=8;
    mapping=0;
    mode='h';
end

if (nargin == 2) && (length(varargin{2}) == 1)
    error('Input arguments');
end

if (nargin > 2) && (length(varargin{2}) == 1)
    radius=varargin{2};
    neighbors=varargin{3};

    spoints=zeros(neighbors,2);

    % Angle step.
    a = 2*pi/neighbors;

    for i = 1:neighbors
        spoints(i,1) = -radius*sin((i-1)*a);
        spoints(i,2) = radius*cos((i-1)*a);
    end

    if(nargin >= 4)
        mapping=varargin{4};
        if(isstruct(mapping) && mapping.samples ~= neighbors)
            error('Incompatible mapping');
        end
    else
        mapping=0;
    end

    if(nargin >= 5)
        mode=varargin{5};
    else
        mode='h';
    end
end

if (nargin > 1) && (length(varargin{2}) > 1)
    spoints=varargin{2};
    neighbors=size(spoints,1);

    if(nargin >= 3)
        mapping=varargin{3};
        if(isstruct(mapping) && mapping.samples ~= neighbors)
            error('Incompatible mapping');
        end
    else
```

```
        mapping=0;
    end

    if(nargin >= 4)
        mode=varargin{4};
    else
        mode='h';
    end
end

% Determine the dimensions of the input image.
[ysize xsize] = size(image);
miny=min(spoints(:,1));
maxy=max(spoints(:,1));
minx=min(spoints(:,2));
maxx=max(spoints(:,2));

% Block size, each LBP code is computed within a block of size
bsizey*bsizex
bsizey=ceil(max(maxy,0))-floor(min(miny,0))+1;
bsizex=ceil(max(maxx,0))-floor(min(minx,0))+1;

% Coordinates of origin (0,0) in the block
origy=1-floor(min(miny,0));
origx=1-floor(min(minx,0));

% Minimum allowed size for the input image depends
% on the radius of the used LBP operator.
if(xsize < bsizex || ysize < bsizey)
  error('Too small input image. Should be at least (2*radius+1) x
(2*radius+1)');
end

% Calculate dx and dy;
dx = xsize - bsizex;
dy = ysize - bsizey;

% Fill the center pixel matrix C.
C = image(origy:origy+dy,origx:origx+dx);
d_C = double(C);

bins = 2^neighbors;

% Initialize the result matrix with zeros.
CLBP_S=zeros(dy+1,dx+1);
CLBP_M=zeros(dy+1,dx+1);
CLBP_C=zeros(dy+1,dx+1);

%Compute the LBP code image

for i = 1:neighbors
  y = spoints(i,1)+origy;
  x = spoints(i,2)+origx;
  % Calculate floors, ceils and rounds for the x and y.
  fy = floor(y); cy = ceil(y); ry = round(y);
  fx = floor(x); cx = ceil(x); rx = round(x);
  % Check if interpolation is needed.
  if (abs(x - rx) < 1e-6) && (abs(y - ry) < 1e-6)
    % Interpolation is not needed, use original datatypes
```

```
    N = d_image(ry:ry+dy,rx:rx+dx);
    D{i} = N >= d_C;
    Diff{i} = abs(N-d_C);
    MeanDiff(i) = mean(mean(Diff{i}));
  else
    % Interpolation needed, use double type images
    ty = y - fy;
    tx = x - fx;

    % Calculate the interpolation weights.
    w1 = (1 - tx) * (1 - ty);
    w2 =      tx  * (1 - ty);
    w3 = (1 - tx) *      ty ;
    w4 =      tx  *      ty ;
    % Compute interpolated pixel values
    N = w1*d_image(fy:fy+dy,fx:fx+dx) +
w2*d_image(fy:fy+dy,cx:cx+dx) + ...
        w3*d_image(cy:cy+dy,fx:fx+dx) +
w4*d_image(cy:cy+dy,cx:cx+dx);
    D{i} = N >= d_C;
    Diff{i} = abs(N-d_C);
    MeanDiff(i) = mean(mean(Diff{i}));
  end
end
% Difference threshold for CLBP_M
DiffThreshold = mean(MeanDiff);
% compute CLBP_S and CLBP_M
for i=1:neighbors
  % Update the result matrix.
  v = 2^(i-1);
  CLBP_S = CLBP_S + v*D{i};
  CLBP_M = CLBP_M + v*(Diff{i}>=DiffThreshold);
end
% CLBP_C
CLBP_C = d_C>=mean(d_image(:));

%Apply mapping if it is defined
if isstruct(mapping)
    bins = mapping.num;
    sizarray = size(CLBP_S);
    CLBP_S = CLBP_S(:);
    CLBP_M = CLBP_M(:);
    CLBP_S = mapping.table(CLBP_S+1);
    CLBP_M = mapping.table(CLBP_M+1);
    CLBP_S = reshape(CLBP_S,sizarray);
    CLBP_M = reshape(CLBP_M,sizarray);
end

if (strcmp(mode,'h') || strcmp(mode,'hist') || strcmp(mode,'nh'))
    % Return with LBP histogram if mode equals 'hist'.
    CLBP_S=hist(CLBP_S(:),0:(bins-1));
    CLBP_M=hist(CLBP_M(:),0:(bins-1));
    if (strcmp(mode,'nh'))
        CLBP_S=CLBP_S/sum(CLBP_S);
        CLBP_M=CLBP_M/sum(CLBP_M);
    end
else
end
```

PART 3 (getmapping.m)

```
function mapping = getmapping(samples,mappingtype)
table = 0:2^samples-1;
newMax  = 0; %number of patterns in the resulting LBP code
index   = 0;

if strcmp(mappingtype,'u2') %Uniform 2
  newMax = samples*(samples-1) + 3;
  for i = 0:2^samples-1
    j = bitset(bitshift(i,1,samples),1,bitget(i,samples));
    numt = sum(bitget(bitxor(i,j),1:samples));
    if numt <= 2
      table(i+1) = index;
      index = index + 1;
    else
      table(i+1) = newMax - 1;
    end
  end
end

if strcmp(mappingtype,'ri') %Rotation invariant
  tmpMap = zeros(2^samples,1) - 1;
  for i = 0:2^samples-1
    rm = i;
    r  = i;
    for j = 1:samples-1
      r = bitset(bitshift(r,1,samples),1,bitget(r,samples));
      if r < rm
        rm = r;
      end
    end
    if tmpMap(rm+1) < 0
      tmpMap(rm+1) = newMax;
      newMax = newMax + 1;
    end
    table(i+1) = tmpMap(rm+1);
  end
end

if strcmp(mappingtype,'riu2') %Uniform & Rotation invariant
  newMax = samples + 2;
  for i = 0:2^samples - 1
    j = bitset(bitshift(i,1,samples),1,bitget(i,samples));
    numt = sum(bitget(bitxor(i,j),1:samples));
    if numt <= 2
      table(i+1) = sum(bitget(i,1:samples));
    else
      table(i+1) = samples+1;
    end
  end
end

mapping.table=table;
mapping.samples=samples;
mapping.num=newMax;
```

PART 4 (TrainTest.m)

```
%% Training And Testing
% assume the number of training images is 6 for both classes.
tic
clear all
clc
N = 6;
Mitotic_SH = [];
Mitotic_MH = [];
Nonmitotic_SH = [];
Nonmitotic_MH = [];
for i = 1 : N
    Name1 = strcat(pwd,'\Pmitotic\cent',int2str(i),'.png');
    eval('Mitotic = imread(Name1);');
    Name2 = strcat(pwd,'\PNonmitotic\cent',int2str(i),'.png');
    eval('Nonmitotic = imread(Name2);');
    [SH MH] = Preprocessing(Mitotic);
    Mitotic_SH = [Mitotic_SH;SH];
    Mitotic_MH = [Mitotic_MH;MH];
    [SH MH] = Preprocessing(Nonmitotic);
    Nonmitotic_SH = [Nonmitotic_SH;SH];
    Nonmitotic_MH = [Nonmitotic_MH;MH];
end
close all;
%% KNN Classifier
%We will use 5 Nearest Neighbors
neighbors = 3;
%read input image to be tested
%%%histogram = [];
for In = 7 : 106
    %%clc;
    Test = strcat(pwd,'\Ptestingmitotic\cent',int2str(In),'.png');
    eval('TestImg = imread(Test);');
    [Test_SH Test_MH] = Preprocessing(TestImg);
    ED_SH = [];
    ED_MH = [];
    %calculate Euclidean Distance between the Mitotic sub-set and
the testing image
    %1- for the S histograms & M histograms
    for j = 1 : N
        ED = sum(abs(Test_SH - Mitotic_SH(j,:)));
        ED_SH = [ED_SH ED];
        ED = sum(abs(Test_MH - Mitotic_MH(j,:)));
        ED_MH = [ED_MH ED];
    end
    %for the Nonmitotic sub-set and the testing image
    %2- for the S histograms & M histograms
    for j = 1 : N
        ED = sum(abs(Test_SH - Nonmitotic_SH(j,:)));
        ED_SH = [ED_SH ED];
        ED = sum(abs(Test_MH - Nonmitotic_MH(j,:)));
        ED_MH = [ED_MH ED];

    end
    %Sort the distances
    Sorted_SH = sort(ED_SH);
    Sorted_MH = sort(ED_MH);

    %find the minimum 5 distances corresponding to 5 neighbors
    A = ismember(ED_SH,Sorted_SH(1,1:neighbors));
```

```
    B = ismember(ED_MH,Sorted_MH(1,1:neighbors));

    %%voting and final decision
    %1-for S histograms
    S_NumberOfNeighborsSayMitotic = sum(A(1:N));
    S_NumberOfNeighborsSayNonmitotic = sum(A(N+1:length(A)));

    disp (Test );
    if S_NumberOfNeighborsSayMitotic >
S_NumberOfNeighborsSayNonmitotic
       disp('S Histograms says: Mitotic')
    else
        disp('S Histograms says: Nonmitotic')
    end
    %2-for H histograms
    H_NumberOfNeighborsSayMitotic = sum(B(1:N));
    H_NumberOfNeighborsSayNonmitotic = sum(B(N+1:length(B)));
    if H_NumberOfNeighborsSayMitotic >
H_NumberOfNeighborsSayNonmitotic
        disp('H Histograms says: Mitotic')
    else
        disp('H Histograms says: Nonmitotic')
    %%%histogram = [histogram; Test_MH Test_SH]
    end
end
toc
```