ANALYSIS OF RED BLOOD CELL (RBC) CLASSIFICATION USING NI VISION BUILDER AI

JALIL BIN LIAS

A project report submitted in partial
fulfilment of the requirement for the award of the
Master of Electrical Engineering

Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia

JUNE 2015

ABSTRACT

Red blood cell (RBC) diagnosis is very important process for early detection of related disease such as malaria and anaemia before suitable follow up treatment can be proceed. Conventional method under blood smears RBC diagnosis is applying light microscope conducted by pathologist. Red blood cell counting and classification only rely on the manual visual inspection which is laborious, tedious and required highly skill and experience pathologist to analyse the shape of the red blood cell. In this project an automated RBC counting and classification system is proposed to speed up the time consumption and to reduce the potential of the wrongly identified RBC. Initially the RBC goes for image pre-processing which involved global threshold of method applied green channel colour image. Then it continues with RBC counting by using particle area and calculator numeric function method. Eventually, Heywood Circularity Factor, Nearest Neighbour, k-Nearest Neighbour and Minimum Mean Distance classifier methods are applied for normal, abnormal and overlap RBC classification. The proposed method has been tested on blood cell images and the effectiveness and reliability of each of the classifier system has been demonstrated.

ABSTRAK

Diagnosis sel darah merah adalah suatu proses yang penting dalam pengesanan awal penyakit yang berkaitan seperti malaria dan anemia sebelum sebarang rawatan lanjutan boleh dilaksanakan. Cara yang biasa digunakan pada ketika ini adalah dengan menggunakan kaedah pengesanan sampel darah di bawah mikroskop cahaya yang dilakukan oleh ahli patologi. Ianya hanya bergantung kepada pengesanan secara visual yang agak rumit dan hanya boleh dijalankan oleh patologi yang berpengalaman sahaja. Dalam projek ini, pengkelasifikasian dan pengiraan sel darah merah dilakukan secara automatik. Cara ini dapat mempercepatkan proses dan mengurangkan tahap kesilapan dalam mengenal pasti bentuk sel darah merah sama ada yang normal atau pun tidak. Ianya bermula dengan imej pra-pemprosesan yang melibatkan proses ‘global threshold’ pada bahagian warna hijau imej. Selepas itu proses diteruskan dengan pengiraan sel darah merah menggunakan cara ‘particle area’ dan ‘numeric function’. Akhirnya pengkelasifikasian sel darah merah dilakukan dengan menngunakan beberapa jenis sistem pengkelasifikasi termasuk ‘Heywood circularity factor’, ‘Nearest Neighbour’, ‘k-Nearest Neighbour’ dan akhir sekali ‘Minimum Mean Distance’ digunakan untuk mengenalpasti sel darah merah yang normal, tidak normal dan juga yang bertindih. Seterusnya, kesemua sistem pengkelasifikasi ini diuji keberkesanan prestasinya secara verifikasi.

CONTENTS

TITLE		**i**
DECLARATION		**ii**
ACKNOWLEDGEMENT		**iii**
ABSTRACT		**iv**
ABSTRAK		**v**
CONTENTS		**vi**
LIST OF TABLES		**ix**
LIST OF FIGURES		**x**
LIST OF SYMBOLS AND ABBREVIATIONS		**xii**
LIST OF APPENDICES		**xiii**
CHAPTER 1 INTRODUCTION		**1**
1.1	Project background	1
1.2	Problem statement	2
1.3	Aim and objective	3
1.4	Scope of works	4
1.5	Outline of thesis	4
CHAPTER 2 LITERATURE REVIEW		**6**
2.1	Introduction	6
2.2	Image acquisition and enhancement	7
2.3	Image conversion	7

2.4	Cell detection	8
2.5	Morphological operation	8
2.6	Image segmentation	9
2.7	Image classification	11
2.8	Summary	11
CHAPTER 3	**METHODOLOGY**	**13**
3.1	Introduction	13
3.2	Acquisition	14
3.3	Pre-processing	14
3.3.1	Global thresholding	15
3.3.2	Global colour thresholding	15
3.3.3	Binary image	17
3.3.4	Morphological operation	17
3.3.5	Remove border object	18
3.3.6.	Erosion	18
3.3.6	Dilation	19
3.4	Heywood circularity factor	20
3.5	NI Vision Builder AI classification function	20
3.5.1	Nearest Neighbour classifier	20
3.5.2	k-Nearest Neighbour classifier	22
3.5.3	Minimum Mean Distance classifier	22
3.5.4	Distance metrics	23
3.6	Summary	25
CHAPTER 4	**RESULT AND ANALYSIS**	**26**
4.1	Introduction	26

	4.2	RBC image pre-processing	26
	4.3	Early studies of Heywood Circularity factor	31
	4.4	RBC features extraction	34
	4.5	RBC classification result	35
	4.6	Classifier performance evaluation	39
	4.7	RBC classification using object classifier function	40
	4.8	Classifier performance evaluation	46
CHAPTER 5	**CONCLUSION**		**53**
	5.1	Conclusion	53
	5.2	Future work	55
REFERENCES			**57**
APPENDICES			**61**

LIST OF TABLES

3.1	Distance metrics for classification methods	29
4.1	Heywood Circularity class of range for each object	32
4.2	Confusion matrix of classes	33
4.3	True or false positive and negative	33
4.4	Classifier performance result	34
4.5	RBC classification result	40
4.6	Image RBC_1 classifier performance result	47
4.7	Classifier performance summary result	47
4.8	RBC images classification average summary result	51

LIST OF FIGURES

1.1	RBCs, WBCs and platelet image in blood sample	3
3.1	Research framework	13
3.2	Flow of acquisition process from microscope (adapted from [21])	14
3.3	RGB histogram average value (a) RBC, (b) RBC background	16
3.4	(a) RGB image, (b) Binary image	17
3.5	(a) Before remove border object, (b) After remove border object	18
3.6	(a) Binary image; (b) Eroded binary image	19
3.7	(a) Before fill hole, (b) After fill hole	19
3.8	Nearest Neighbour	21
3.9	k-Nearest Neighbour	22
3.10	Minimum Mean Distance	23
3.11	Comparison of metrics in the value of 1 unit distance	25
4.1	Image pre-processing flow diagram	27
4.2	a) RBC_1, (b) RBC_2, (c) RBC_3 and (d) RBC_4	28
4.3	Green colour channel manual threshold process	28
4.4	Vision Assistant process flow process	29
4.5	(a) Threshold, (b) After threshold, (c) Reverse, (d) Remove border object, (e) Remove small object, (f) Fill hole, (g) Equalize and (h) Reverse.	30
4.6	Samples of items	
4.7	Images classification result	32
4.8	RBC classification flow diagram.	36
4.9	(a) Non-overlap RBC particle area range setting, (b) Normal RBC Heywood circularity factor range setting.	37
4.10	Before and after classification result of (a) RBC_1, (b) RBC_2,	

	(c) RBC_3 and (d) RBC_4.	38
4.11	RBC Classification flow chart	42
4.12	Image training stage	43
4.13	(a) Classifier trained image dataset (b) Classifier testing process	44
4.14	Image RBC_4 classification result	45
4.15	Ground truth data of RBC classification; (a) RBC_1, (b) RBC_2, (c) RBC_3, (d) RBC_4	46
4.16	Image RBC_1 classifier accuracy result	48
4.17	Image RBC_2 classifier accuracy result	48
4.18	Image RBC_3 classifier accuracy result	49
4.19	Image RBC_4 classifier accuracy result	49
4.20	Overall classifier accuracy result	50
4.21	Classifier performance evaluation comparison result	50
4.22	RBC images comparison result	52

LIST OF SYMBOLS AND ABBREVIATIONS

RBC	-	Red blood cell
WBC	-	White blood cell
NI	-	National Instrument
AI	-	Automated inspection
PCNN	-	Pulse couple neural network
ANN	-	Artificial neural network
WOSRAS	-	Wireless object sorting robot arm system
CCL	-	Connected component labeling
CHT	-	Circular Hough transform
SEM	-	Scanning electron microscope
SE	-	Structuring element
HT	-	Hough transform
HCF	-	Heywood circularity factor
NN	-	Nearest Neighbour
k-NN	-	k- Nearest Neighbour
MMD	-	Minimum Mean Distance
TP	-	True positive
FP	-	False positive
TN	-	True negative
FN	-	False negative
LPG	-	Long pitch gear
SPG	-	Short pitch gear
RGB	-	Red green blue

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A1-A18	Classifier performance result	61-78
B	Nearest Neighbour with Sum Metric result	79
C	Classifier result	80
D1-D3	Classifier summary result	81-83
E1-E4	RBC classification summary result	84-87
F1-F7	Conference paper	88-97

CHAPTER 1

INTRODUCTION

1.1 Project Background

Blood is determined by connective tissue that consisting of cells suspended in plasma [1]. Its major function is to convey various agents including oxygen, carbon dioxide, nutrients, wastes, and hormones. Blood cells are composed of erythrocytes (red blood cells, RBCs), leukocytes (white blood cells, WBCs) and thrombocytes (platelets). The majority cells abundant in blood are small reddish cells called erythrocytes or red blood cell. An erythrocyte is a discoid cell with a thick rim and a thin sunken center [2]. The main function of RBC is to move oxygen from lung to tissues in body and collect carbon dioxide from tissue back to the lung. Whereby, white blood cell or Leukocytes is the cell part of immune system.

Diagnosis of RBC in medical area contributes information about pathological diseases and condition. The shape and quantity of RBC in samples can be connected to the relevant diseases. Thus, the detail and accurate analysis is important to serve the correct treatment for the patient. The time consumption for the analysis to be undergone will affect to the early treatment process for the patient. Complete blood count analysis (CBC) is a process involving RBC. Any abnormal finding from the result can be signed for disease such as anaemia and secondary effect of several other disorders [3]. Factors that should be consider during performing RBC counting including level of age of people (children and adult, younger and older) and strenuous physical activity [3].

When performing the process, any abnormal and overlapping RBC should be identified prior to count the total number. The former cell should be excluded from the normal cell count while the latter need to be separated before considered for counting. Since such a process is a tedious and time consuming, an image processing is a handy medium for assisting the operator by labelling normal, abnormal and overlapping RBC and provides the numbers of normal and abnormal cell. This will speed up the counting time and the same time reduction the human cause error.

Image processing is powerful method to identify each single cell in blood samples. Compare to conventional method by using haemocytometer, image processing provide advantages that really helpful in RBC analysis. Classification of each cell in blood samples is the ultimate process to be conducted to identify each single cell in the blood cell. This classification process will lead to the counting process. From counting process the total quantity of each single cell in blood samples can be gained and consequently provide a conclusion about the health status of the patient.

1.2 Problem Statement

The conventional device used to count blood cell is the haemocytometer. It consists of a thick glass microscope slide with a rectangular indentation creating a chamber of certain dimensions. This chamber is etched with a grid of perpendicular lines. It is possible to count the chamber of cells in a specific volume of fluid and calculate the concentration of cells in the fluid [3][4]. To count blood cell, physician must view haemocytometer through a microscope and count blood cells using hand tally counter. The overlapped blood cells can’t be counted by using haemocytometer (Figure 1.1). Furthermore, other cells besides normal RBC such as WBC, irregular shape of RBC and platelet are also elements that interfere during RBC counting.

Normally, the counting task is time-consuming and laborious. Furthermore, conventional method is considering time-consuming to complete the counting task and it is laborious [5][7]. Before this counting process can be conducted, the identification of each single cell in the blood sample needs to be done. Identification by human is a current practice in this process. Experience and knowledge will help

this identification process which lead to RBC counting process can be conducted in short time consumption. Development of automated system that can identify or classify each of single cells in the blood samples will help to overcome the burden of a manual process. This automated system will really helpful for haematologist or medical practitioner.

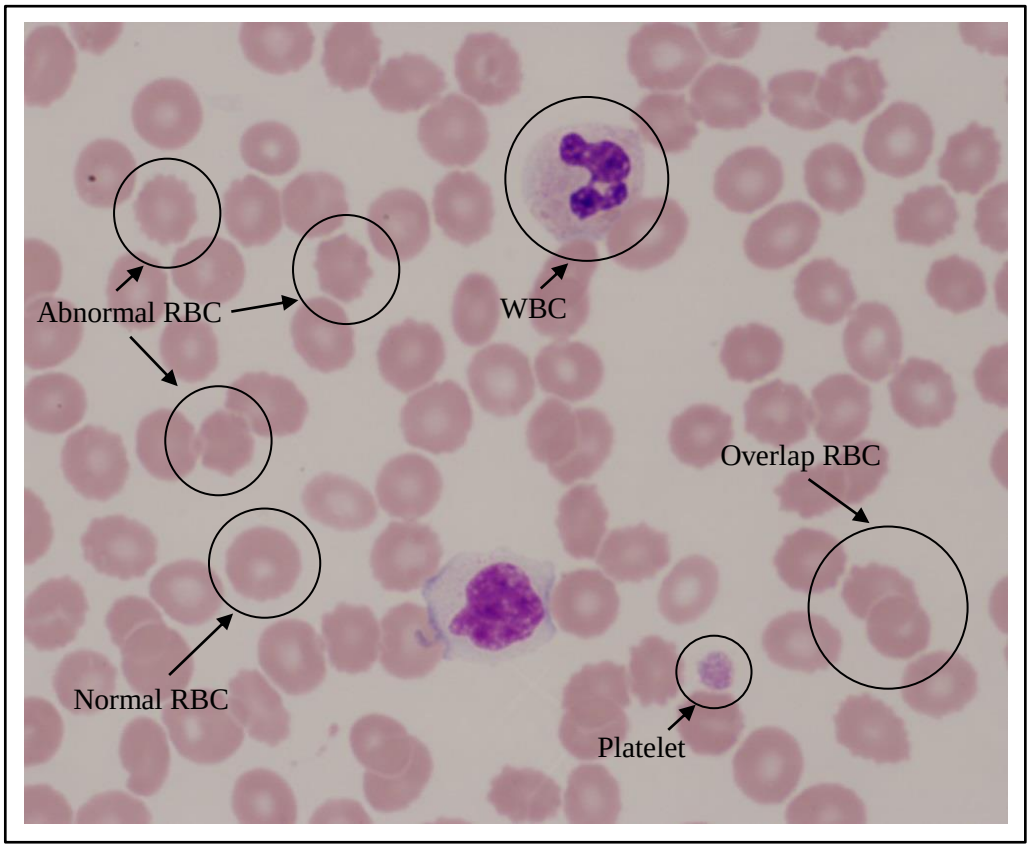

Figure 1.1: RBCs, WBCs and platelet image in blood sample

1.3 Aim & Objectives

The project aim is to classify the RBC into normal and abnormal in an overlapping condition before counting the number of each individual cluster. To achieve the aim of this research, the following objectives are formulated:

i. To develop a method for automatically segment out RBC region.
ii. To classify the normal RBC and irregular/abnormal RBC from blood smear images.

iii. To assess the performance of the develop system qualitatively and quantitatively.

1.4 Scope of works

i. Develop RBC image pre-processing.
ii. Develop RBC image classifier system.
iii. Produce the quantity of RBC normal, abnormal and overlap after the classification process.
iv. Evaluate the performance each of the RBC classification system.

1.5 Outline of thesis

This project is classified into five chapters. The scope of each chapter is explained as below:

First chapter gives the background of the thesis, problem statement, aim and objective, scopes of works and outline of the thesis.

Chapter II is about the literature review, in which previous studies and theories related to this project are discussed and reviewed. It is also describe about RBC image classification using several methods such as connected component labelling, neural network and nearest mean. Literature review provides a background of this project and also gives and direction in this research.

Chapter III deals with a research methodology. It describes the detailed methods that have been used to conduct this project. This chapter proposes the method that involved in this project including image pre-processing using morphological and threshold method. The classification method that is proposed is Heywood circularity factor, Nearest Neighbour, k-Nearest Neighbour and Minimum Mean Distance.

Chapter IV is for the results and discussion. This chapter will highlight the result of each classifier method that is proposed in this project also the each of the performance evaluation conducted to find the most suitable classifier that provides by NI Vision Builder AI that suits with this RBC classification project.

Chapter V concludes this project. It also describes the next step that need to be done in the future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

RBC analysis through image processing method has been subject of interest of many researchers recently because of conventional method using haemocytometer is quite tedious and time consuming. Most of the previous work using MATLAB image processing toolbox as their main tool for analyse the RBC image because of its convenient for evaluating newly developed algorithm.

Many combination of method has been tested in the RBC image processing. The main challenge of such process is identification of normal RBC in the blood sample since the variety in shape of such cell may exist. It will become worse if the cell is clump and overlap in a group. Such a problem is a main challenge for a researcher to identify or to classify single RBC in this region.

Some of the previous works just ignore the overlapped RBC [8][9] and some of it proposed a method to classify/separate the overlapped cell [7][10][11] by using sophisticated image processing methodology. In this project, the focus is on an automated method for processing the RBC under various overlapping condition. In following section, explanation of various methods available to overcome such an issue is given in detail.

2.2 Image Acquisition and Enhancement

The images basically are obtained from the samples of blood that being captured using light microscope which the process involve the preparation of blood smear [8][12]. Blood smear is a process to put the blood specimen on the slide that being observed under the microscope. The images are then being filtered to reduce or minimize noise. Several filtration techniques are used such as average filter and median filter [8][13][14]. Average filter is a linear spatial domain filter and function to decrease all noises in sample. It uses a defined filter mask to average grey level pixel in the neighbourhood.

While median filter is a nonlinear spatial filter that changes the gray value at the center pixel with median value of the gray value of the pixel group. Edge detection reduces amount of data, filters useless information and preserves important structural details. Histogram equalization is used to adjust intensity value of image [15][10][16]. Contrast and brightness adjustment is a step that has been used in image processing. Both adjustments used histogram of interested image to display the range of intensity value of image [10].

2.3 Image Conversion

Illumination issue always happened in blood cell microscopic image. To avoid illumination issue, color conversion method is applied. Previous study on color conversion is done by using Ycbcr technique where the RGB color is divided into Y, cb and cr component to avoid illumination [8]. The second component of the Ycbcr color has been chosen and it shows the clear appearance of the WBC nucleus and platelet. Then, color and contrast and brightness adjustment method has been done in the next process to give the view of color representation image.

Previous method has been done on image conversion from original image to gray image [17][18][13][19]. Classifying the image by gray-level pixels may reduce and simplify some image processing operations such as edge detection, edge smoothing, feature extraction, image processing and image registration [15].

RGB image conversion into binary image also has been done on previous work [20][30]. Conversion to binary image is helpful to identify foreground and background of the image. Due to the binary image is in black and white mode, thus the image can be easily identified. This process will be continued by threshold method, threshold will be used as reference to identify object and its background.

2.4 Cell Detection

Cell detection is one of the methods to identify the perimeter or boundary cell. One of the most popular cell detection that been used is edge detection method. Most edge detection methods such as Canny or Sobel edge detectors exploit the image intensity gradient magnitude to identify object boundaries in image [9][10]. Edge detection does not work well between two overlapping cell. This is due to the change of intensity between two overlapping cells is very slow. That's why it not suitable for detection of inter-cell-boundaries.

2.5 Morphological Operation

Morphology is a broad set of image processing operations that process images based on shapes [19]. Morphological operations apply a structuring element to an input image, creating an output image of the same size. Morphological image processing is based on a strong mathematical concept which been used to change the size, shape, structure and connectivity of objects in the image [8].

In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image [10][30]. This adding and removing object is also called 'dilation' and 'erosion'.

Morphological operation is used to separate overlapped image. But there is only certain condition where morphological operation can be used. It is normally works to separate minor overlapped image. This method is not suitable to separate over overlapped image.

All morphological processing operations are based on two simple ideas, hit and fit. Fit stands for the condition when all pixels in the structuring element cover on pixels in the image whereas hit signifies the condition when any of the pixels in the structuring element covers on a pixel in the image.

Morphological operators also include a few steps which are filling holes, area calculation, template calculation, opening, closing and reconstruction. Mathematical morphological operators used to segment RBC by eliminating WBC appearance [8] [18]. It is used for extracting image components and useful in represent or describe the region of shape such as boundary, skeleton and texture [16].

2.6 Image Segmentation.

Segmentation is one of the most crucial tasks in image processing and computer vision [21]. As mentioned earlier blood cell contains RBC, WBC, platelet and sickle RBC. To identify each of this item, there are several method of image processing has been done. Once each item can be segmented, the analysis can be done separately.

From previous work that done, the most challenge scenario in RBC image segmentation identifying single cell in overlapped condition. However, many researches did the improvement by combining several methods or create the new technique to overcome the problem.

One of the methods for image segmentation is watershed transform. Watershed is a morphological technique that derives its name from an expression in geography, where watershed is defined as the ridge that divides areas drained by different river systems [21]. Watershed algorithm is a method used to segment RBC in overlapping area [8][9][19]. However, it cannot handle when the overlapping area contain important information and it is hard to ensure the accuracy of segmentation due to the large error. The improvement has been done by combining mathematical morphology using corrosion and expansion algorithm with the principle of watershed algorithm [7].

The distance transform is a useful tool employed in conjunction with the watershed transform. It computes the distance from every pixel to the nearest non-zero valued pixel. On previous work, distance transform is used combined with

watershed for splitting clumped cells. The main function of distance transform is it detects the cell central point. Thus if the image is inconsistent in shape or overlapped, by using this method it can detect cell image based on the central point's [17].

The other recent common methods used for overlapping and clumped cells are concavity analysis and template matching [8]. On the other hand, concavity analysis is used to measure split lines for an overlapped cells. Nevertheless, it is only applicable for a pair of cells but useless against multiple overlapping cells. Plus, a very accurate segmentation is needed to apply this method. Other technique template matching which uses a template of RBC or clumping area to be matched to the object in the image able to separate small cell in shape and size. However, it is computationally expensive.

In a research, template matching method was combined with pulse coupled neural network (PCNN) since PCNN cannot cope with overlapped cells. However the accuracy decreases whenever the RBCs are overlapped totally because the area of one cell is considered as a template and the algorithm works only in 100x microscope scale [18].

Hough transform method also been used in previous work. The Hough Transform (HT) has been recognized as a very powerful tool for the detection of parametric curves in images [11][13][22]. It implements a voting process that maps image edge points into manifolds in an appropriately defined parameter space. The Circular Hough Transform (CHT) is one of the modified versions of the HT. The CHT concentrates to find circular patterns within an image. The Circle Hough Transform is designed to find a circle characterized by a center point.

Contour tracing approach has been used to segment scanning electron microscope (SEM) images. The method views contour detection and negotiating perceived problem areas one at a time but it still has lack when facing overlapping cells. It applies Bayesian tracking framework [1]. Consequently, the RBC segmentation of SEM image utilized shape reconstruction and multi-scale surface based on shape from shading technique combined with linear approximation [14]. Other than that, classification of RBC has been done through depth map and surface feature for different surface shapes [23].

2.7 Image Classification

In image classification, Multilayer perceptron artificial neural network (ANN) is most common method used to identify and count RBC [15][10]. It performed by adjusting the value of the weight between the elements and the mean square error is calculated from there. The weight value is used to separate abnormal RBC and Normal RBC, thus the quantity of normal RBC is counted. The result shows 74% accuracy of counting RBC [10]. In other research ANN used in observing the relation between Hgb level and RBC with the color or blood sample. More than 90% of the samples is produced with tolerance less than 5% compared to lab results [12]. In some other research, back projection of ANN compared with connected-component labelling (CLL) has been proved. Haematology analyser Sysmex KX-21 was used as benchmark for the comparison. The average accuracy of the CCL is 87.74% and the back projection ANN produced 86.97% of accuracy [5].

A comparative study among Nearest Neighbour, k-Nearest Neighbour and Minimum Mean Distance classifier of NI Vision Builder AI has been conducted in Wireless Object Sorting Robot Arm System (WOSRAS) [29]. The result from the classification shows that the Nearest Neighbour with Sum metric distance shows the highest performance in term of accuracy, misclassification rate and Kappa-coefficient.

2.8 Summary

RBC classification using image processing has been done in many previous works. As we know RBC analysis using image processing is not a new thing in medical diagnosis. Researchers focus on the improvement of the accuracy and promising result in their research by using many different methods. There is a challenge in machine vision system to achieve the quality level of human vision system.

There are still weaknesses and constraints due to the image itself such as color similarity, weak edge boundary, overlapping condition, image quality, contrast, brightness, illumination and noise. Thus, more study must be done to handle those matters to produce strong analysis approach for medical diagnosis purpose. This

project is hoped can build a better solution and help to improve the current methods so that it can be more capable, robust, and effective whenever any sample of blood cell is analysed.

CHAPTER 3

METHODOLOGY

3.1 Introduction

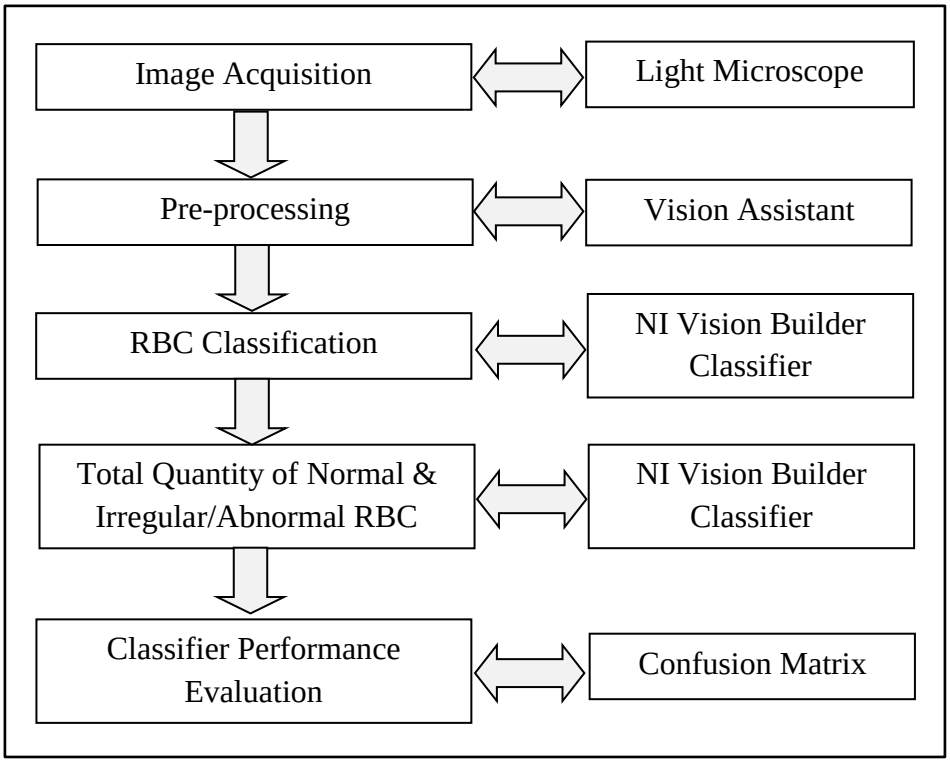

Figure 3.1: Research framework

There are several steps taken in for RBC classification and counting from previous works. The problem domain in this case is to extract the RBC from a blood cell image automatically. Whereby, the goal is to classify RBC between normal and

irregular shape. The foundation methods that are taken in digital image processing will be similar one to another. Image processing is not a one-step process: most solutions follow a sequential processing scheme whose main steps are described next. Figure 3.1 shows the overall flow process of the research.

3.2 Acquisition

The acquisition block is in charge of acquiring one or more images of blood samples of anaemia patient or person that facing RBC disease. This acquisition process will be gained from digital microscope. Several factors should be considered on the blood image that captured from the digital microscope that will likely impact the quality of the result of the RBC classification such as blur and illumination. Figure 3.2 shows the flow of acquisition process of blood cell from microscope.

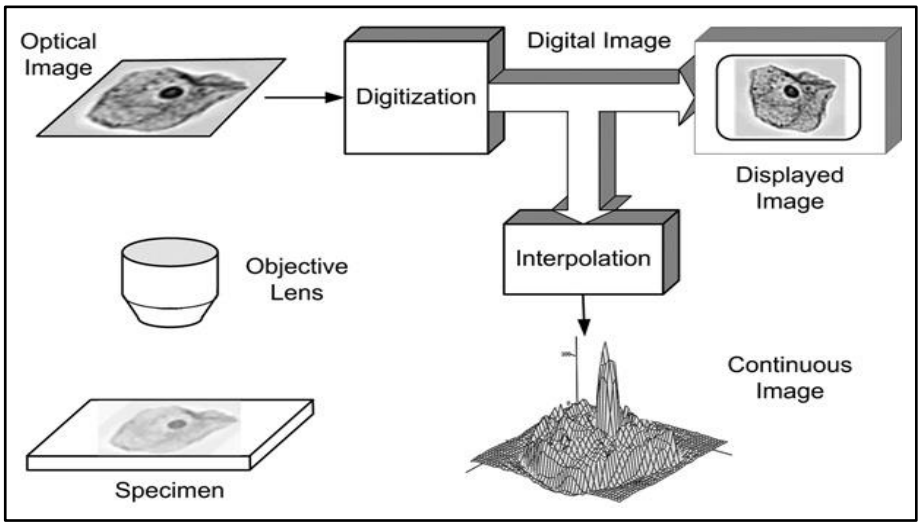

Figure 3.2: Flow of acquisition process from microscope

3.3 Pre-processing

The goal of the pre-processing stage is to improve the quality of the acquired image. Possible algorithms to be employed during this stage include contrast improvement, brightness correction, and noise removal. As described earlier, blood cell contains RBC, WBC and platelet.

During pre-processing, unwanted image need to be removed. This process will remain RBC as the remaining object to be analysed for next process. Possible algorithms to be employed during this stage including border image removal, removing small objects and filling holes of the RBC images.

3.3.1 Global Thresholding

Many optimal strategies for selecting threshold values have been suggested in the literature. These strategies usually rely on assumed statistical models and consist of modelling the thresholding problem as a statistical inference problem. Unfortunately, such statistical models usually cannot take into account important factors such as borders and continuity, shadows, non-uniform reflectance, and other perceptual aspects that would impact a human user making the same decision. Consequently, for most of the cases, manual threshold selection by humans will produce better results than statistical approaches would.

The well-known global image threshold is Otsu's method. Otsu's method is histogram based image thresholding method that separates the image pixel into two classes with a minimal intra-class variance. In this project, global colour thresholding process is applied to convert the image from RGB to binary image.

3.3.2 Global Color Thresholding

Colour thresholding converts a colour image into a binary image. To threshold a colour image, specify a threshold interval for each of the three colour components. A pixel in the output image is set to 1 if and only if its colour components fall within the specified ranges. Otherwise, the pixel value is set to 0.

For a pixel in the colour image to be set to 1 in the binary image, its red value should lie between 130 and 200, its green value should lie between 100 and 150, and its blue value should lie between 55 and 115. Figure 3.3 shows the difference of RGB histogram value of RBC compare to its background.

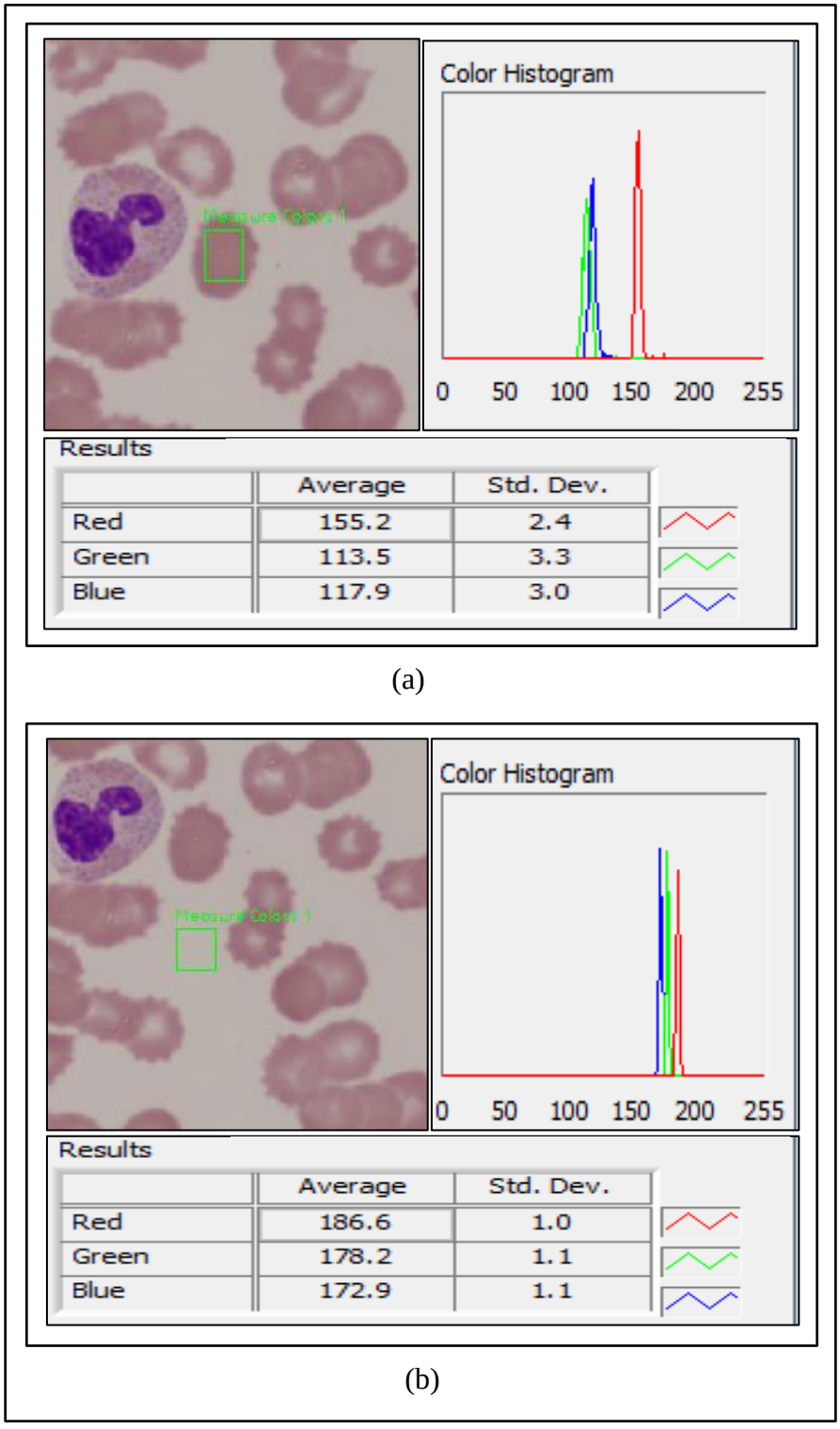

Figure 3.3: RGB histogram average value (a) RBC, (b) RBC background

3.3.3 Binary Image

Binary images are encoded as a 2D array, typically using 1 bit per pixel, where a 0 usually means “black” and a 1 means “white” (although there is no universal agreement on that). The main advantage of this representation usually suitable for images containing simple graphics, text, or line art is its small size. Figure 3.4 shows conversion from RGB to binary image.

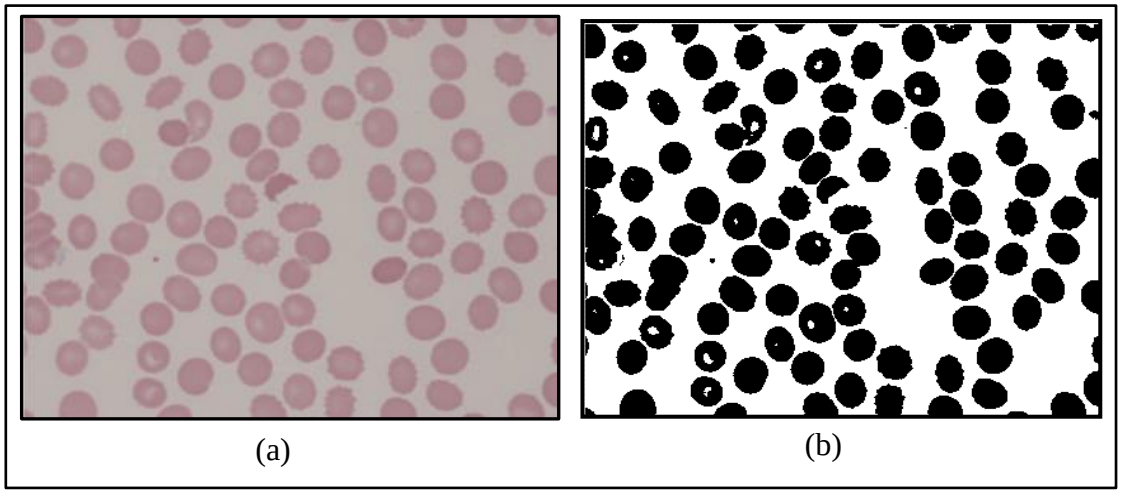

Figure 3.4: (a) RGB image, (b) Binary image

3.3.4 Morphological Operation

Morphology is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours.

The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. All morphological processing operations are based on two simple ideas, hit and fit. Fit stands for the condition when all pixels in the structuring element cover on pixels in the image whereas hit signifies the condition when any of the pixels in the

structuring element covers on a pixel in the image. The different morphological operators used are discussed below.

3.3.5 Remove Border Object

Objects that touching border is not in incomplete shape and it is difficult to be classified. Due to this constrain, the RBC image that touches border can be eliminated. This process can affect the performance result of the classifier but due to the limitation these incomplete shape of RBC need to be remove. Figure 3.5 shows before and after result of this process.

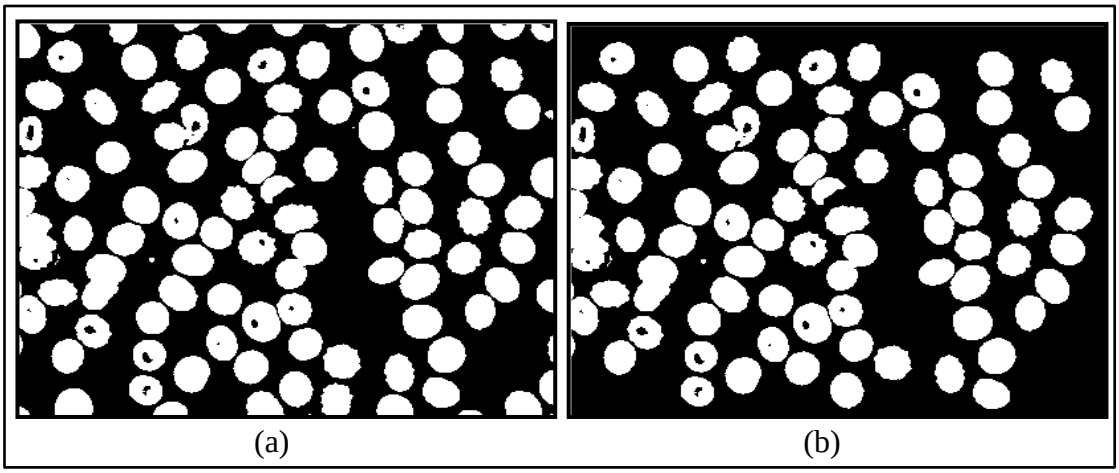

Figure 3.5: (a) Before remove border object, (b) After remove border object

3.3.6 Erosion

Erosion is a morphological operation whose effect is to “shrink” or “thin” objects in a binary image. The direction and extent of this thinning is controlled by the shape and size of the structuring element. In this project, erosion function is used in removing small object as per shown in Figure 3.6.

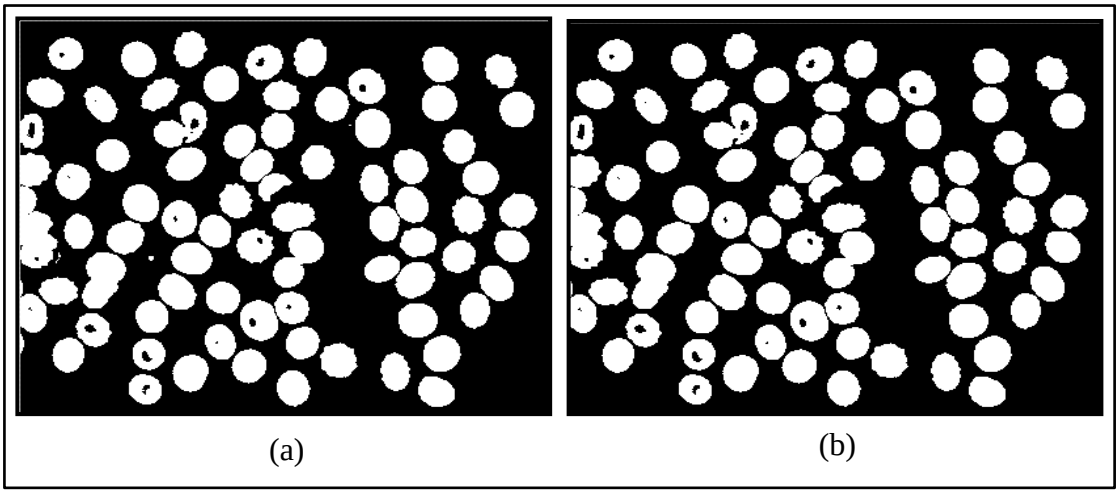

Figure 3.6: (a) Binary image; (b) Eroded binary image

3.3.7 Dilation

Dilation is a morphological operation whose effect is to “grow” or “thicken” objects in a binary image. The extent and direction of this thickening are controlled by the size and shape of the structuring element. The structuring element (SE) is the basic neighbourhood structure associated with morphological image operations. It is usually represented as a small matrix, whose shape and size impact the results of applying a certain morphological operator to an image. In this project, the dilation function is applied in filling the hole in RBC images as per shown in Figure 3.7.

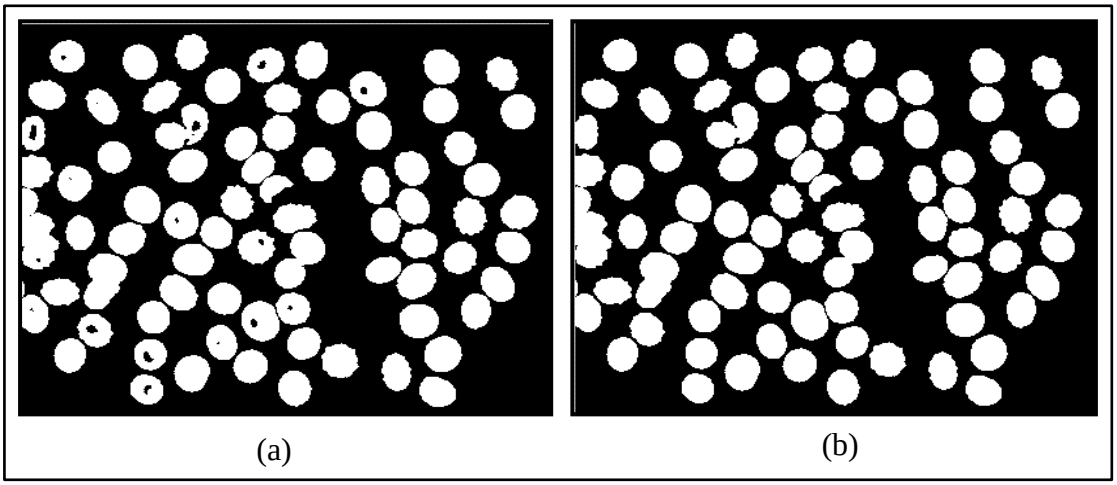

Figure 3.7: (a) Before fill hole, (b) After fill hole

3.4 Heywood Circularity Factor

Heywood circularity factor is used as image classifier inside RBC images to identify non-overlap, overlap and normal RBC. Heywood circularity Factor is perimeter divided by the circumference of a circle with the same area. The closer the shape of a particle is to a disk, the closer the Heywood circularity factor is to 1. Heywood circularity factor is also a ratio of particle perimeter to the perimeter of the circle with the same area. It is defined in (3.1):

$$Heywood\ circularity\ factor = \frac{particle\ perimeter}{perimeter\ of\ circle\ with\ same\ area\ as\ particle} \quad (3.1)$$

$$= \frac{particle\ perimeter}{2\sqrt{\pi\ x\ particle\ area}}$$

3.5 NI Vision Builder AI Classification Function

NI Vision Builder AI provides image classification function that can be used to classify objects in several types of application. There are three classifiers that provided in the system which is Nearest Neighbour (NN), k-Nearest Neighbor (k-NN) and Minimum Mean Distance. Along with these three classifiers, there are three types of metrics that can be selected. The three metrics are Maximum, Sum and Euclidean. In the RBC classification project, the approach of classification will be attempted for those three classifier and the three metrics and the performance of each classifier can be observed.

3.5.1 Nearest Neighbour Classifier

Nearest Neighbour is the most direct approach to classification. In Nearest Neighbour classification, the distance of an input feature vector of unknown class to another class is defined as the distance to the closest samples that are used to represent that class. In Nearest Neighbour (NN) algorithm, the input feature vector *X*

of unknown class to a class C_j is determined as the distance to the nearest neighbour which is used to represent the class as shown in (3.2).

$$d(X, C_j) = {}^{min}_{i} d(X,\ X_i^j) \tag{3.2}$$

Where $d(X, X_i^j)$ is the distance between X and X_i^j

The classification rule assigns a pattern X of unknown class to the class of its nearest neighbour that is given in (3.3).

$$X \in Class\ C_j, if\ d(X, C_j) = {}^{min}_{i} d(X,\ C_i) \tag{3.3}$$

The main advantage of Nearest Neighbour algorithm is its simplicity approach for classification. It works well if corresponding feature vectors for every class are available. Nearest Neighbour classification is the most intuitive approach for classification. If representative feature vectors for each class are available, Nearest Neighbour classification works well in most classification applications. Figure 3.8 shows Nearest Neighbour concept.

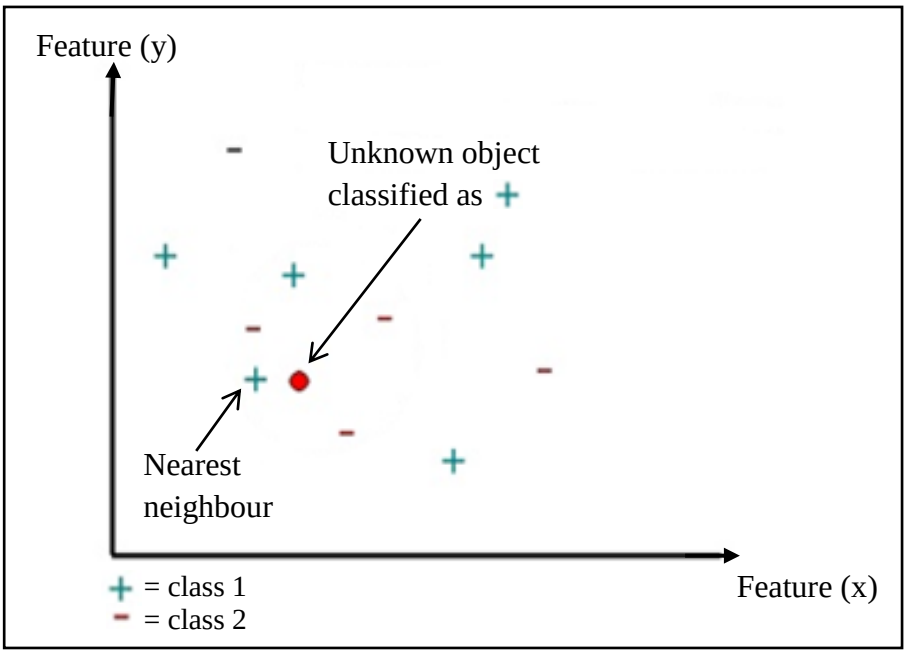

Figure 3.8: Nearest Neighbour

3.5.2 k-Nearest Neighbour Classifier

In k-Nearest neighbour classification, an input feature vector X is classified into a class C_j based on a voting mechanism. The NI Classifier finds k nearest samples from all the classes. The input feature vector of unknown class is assigned to the class with majority of the votes in the k nearest samples.

The outlier feature patterns caused by noise in real-world applications can cause incorrect classifications when Nearest Neighbour classification is used. K-Nearest Neighbour with k=3 is illustrates in Figure 3.9.

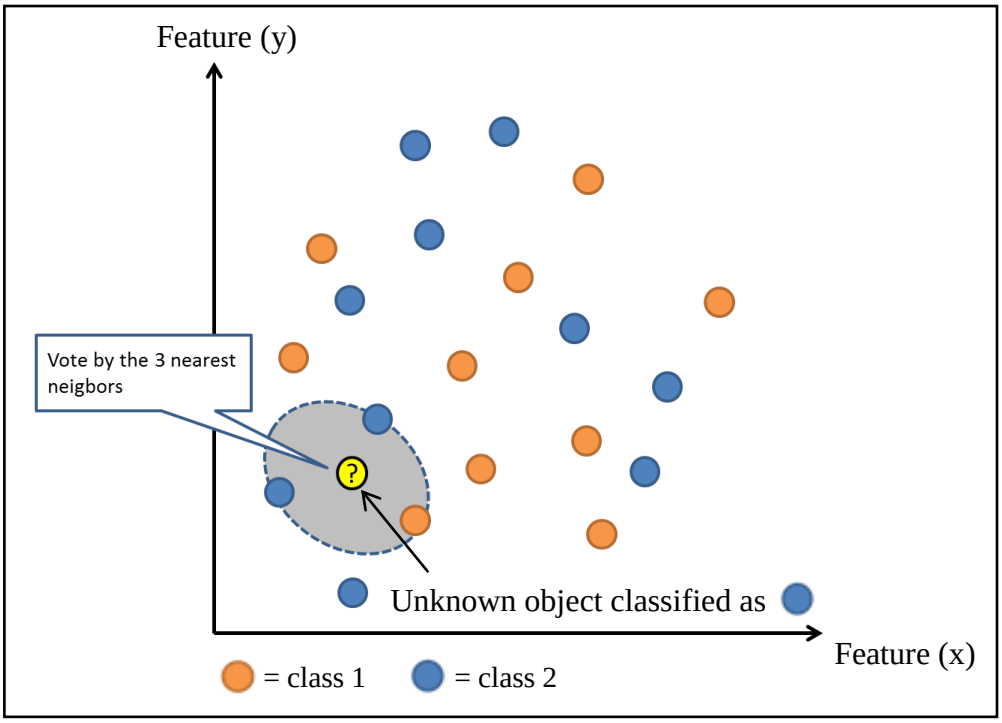

Figure 3.9: k-Nearest Neighbour

3.5.3 Minimum Mean Distance Classifier

The last one is Minimum Mean Distance which is most effective in applications that have little or no feature pattern variability or other corruptive influences. In minimum mean distance classification, an input feature vector of unknown class is classified based on its distance to each class centre. Consider that $\{X_1^j, X_2^j, \ldots, X_{n_j}^j\}$

be n_j feature vectors which is used to represent the class C_j. Every feature vector contains a label of class j that has been selected for representing the class. The centre of the class j is given in (3.4)

$$M_j = \frac{1}{n_j}\sum_{i=1}^{n_j} X_i^j \tag{3.4}$$

An input feature vector X of unknown class was classified in the classification phase depends on the distance to each class centre and given in (3.5)

$$X \in Class\ C_j, if\ d(X, M_j) = \ ^{min}_{i} d(X, M_i) \tag{3.5}$$

Figure 3.10 shows the Minimum Mean Distance Classification process.

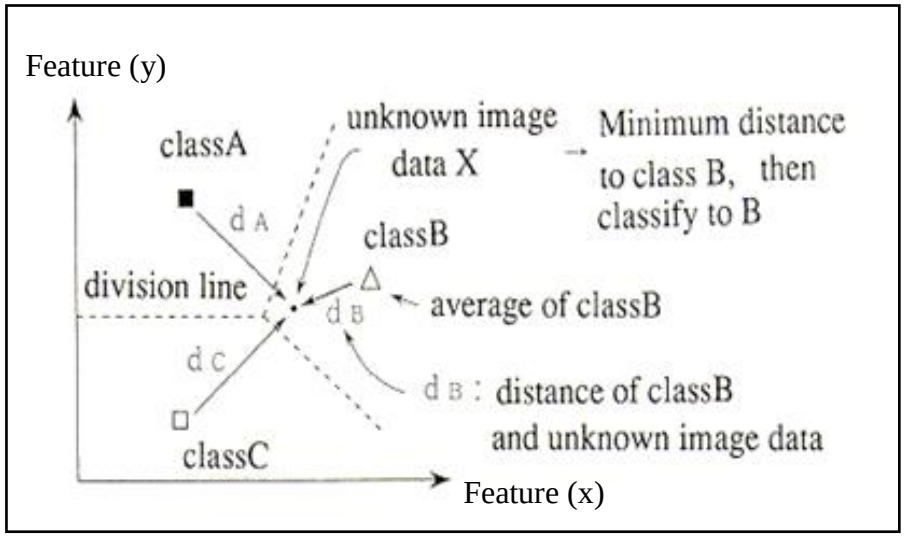

Figure 3.10: Minimum Mean Distance

3.5.4 Distance Metrics

The classifier comes along with metric or distance. There are three metrics that can be selected for each classifier which is Maximum, Sum and Euclidean. Maximum is most sensitive to small variations between samples. Maximum is used to classify samples with very small differences into different classes. While Sum metric is used

in most classification applications. Sum is also known as the Manhattan metric or Taxicab metric. The last one is the Euclidean metric that least sensitive to small variations between samples. Euclidean metric is applied in classification samples with small differences into the same class.

Let $X=[x_1, x_2,...,x_n]$ and $Y=[y_1, y_2,..., y_n]$ be the feature vectors, then the distance metric $d(X,Y)$ was given for each distance metric. Euclidean distance (L2), Sum distance (L1) and Maximum distance (L∞) was evaluated and the resultant formulas for the distance metrics of classification methods were shown in Table 3.1.

Table 3.1: Distance metrics for classification methods

Euclidean distance (L2)	$d(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$
Sum distance, also known as the City-Block metric or Manhattan metric (L1)	$d(X,Y) = \sum_{i=1}^{n}\lvert X_i - Y_i\rvert$
Maximum distance (L∞)	$d(X,Y) = \max_i \lvert X_i - Y_i\rvert$

Figure 3.11 shows the comparison distance for value of 1 unit among the three metrics. It can be seen that the Maximum Mean Distance metric shows square shape, Sum metric shows diamond shape and Euclidean metric shows circle shape. It can be conclude that each metric has different length of distance to reach the same coordinate of object.

REFERENCES

[1] J. Vromen and B. McCane, "Red blood cell segmentation from SEM images," *2009 24th Int. Conf. Image Vis. Comput. New Zealand, IVCNZ 2009 - Conf. Proc.*, pp. 44–49, 2009.

[2] K. S. Saladin, *Antomy and Physiology: The unity of form and function*, no. McGraw-Hill, NY 4th. 2007, p. chap. 18, pp. 680–696.

[3] D. R. Caprette, *, "Introduction Laboratory : using a counting chamber," Bios211, 2007*. 2007, p. 211.

[4] J.P. Mather and P.E. Roberts, *Introduction to cell and tissue culture: Theory and Technique,*. 1998.

[5] a. M. T. Nasution and E. K. Suryaningtyas, "Comparison of red blood cells counting using two algorithms: Connected component labeling and backprojection of artificial neural network," *2008 IEEE PhotonicsGlobal Singapore, IPGC 2008*, 2008.

[6] M. J. Su, Z. Bin Wang, H. J. Zhang, and Y. De Ma, "A new method for blood cell image segmentation and counting based on PCNN and autowave," *2008 3rd Int. Symp. Commun. Control. Signal Process. ISCCSP 2008*, no. 60572011, pp. 6–9, 2008.

[7] J. Huang, "An improved algorithm of overlapping cell division," *Proc. - 2010 Int. Conf. Intell. Comput. Integr. Syst. ICISS2010*, pp. 687–691, 2010.

[8] J. M. Sharif, M. F. Miswan, M. A. Ngadi, S. Hj, and M. Mahadi, “Red Blood Cell Segmentation Using Masking and Watershed Algorithm : A Preliminary Study,” no. February, pp. 27–28, 2012.

[9] Y. Karunakar and A. Kuwadekar, “An unparagoned application for red blood cell counting using marker controlled watershed algorithm for android mobile,” *Int. Conf. Next Gener. Mob. Appl. Serv. Technol.*, pp. 100–104, Sep. 2011.

[10] J. Poomcokrak and C. Neatpisarnvanit, “Red blood cells extraction and counting,” no. Isbme, pp. 199–203, 2008.

[11] B. Venkatalakshmi and K. Thilagavathi, “Automatic Red Blood Cell Counting Using Hough Transform,” no. Ict, pp. 267–271, 2013.

[12] S. Zahir, R. Chowdhury, G. W. Payne, and P. George, “Automated Assessment of Erythrocyte Disorders Using Artificial Neural Network,” pp. 776–780, 2006.

[13] M. A. A. Ra. Nasrul Humaimi Mahmood, Poon Che Lim, Siti Madihah Mazalan, “Blood Cell Extraction Using Color Based Segmentation Technique,” vol. 2, No. 2 A, 2013.

[14] R. Wang, B. MacCane, and B. Fang, “RBC Image Segmentation Based on Shape Reconstruction and Multi-scale Surface Fitting,” *2010 Third Int. Symp. Inf. Sci. Eng.*, pp. 586–589, Dec. 2010.

[15] N. D. Jambhekar, “Red Blood Cells Classification using Image Processing,” vol. 1, no. 3, pp. 151–154, 2011.

[16] C. Di Ruberto, A. Dempster, S. Khanh, and B. Jarratt, “Segmentation of Blood Images Using Morphological Operators t ttNationallnstitute of Medical Research , London , UK,” pp. 397–400, 2000.

[17] N. T. Nguyen, A. D. Duong, and H. Q. Vu, "A new method for splitting clumped cells in red blood images," *Proc. - 2nd Int. Conf. Knowl. Syst. Eng. KSE 2010*, no. 1, pp. 3–8, Oct. 2010.

[18] S. S. Adagale and S. S. Pawar, "Image segmentation using PCNN and template matching for blood cell counting," *2013 IEEE Int. Conf. Comput. Intell. Comput. Res.*, pp. 1–5, 2013.

[19] H. Tulsani, "Segmentation using Morphological Watershed Transformation for Counting Blood Cells," vol. 2, no. Iii, pp. 28–36, 2013.

[20] H. Berge, D. Taylor, S. Krishnan, and T. S. Douglas, "MRC / UCT Medical Imaging Research Unit , Department of Human Biology Division of Clinical Pharmacology University of Cape Town , South Africa," pp. 204–207, 2011.

[21] O. Marques, *Proctical Image and Video Processing Using MATLAB*. 2011.

[22] W. Xiong, S. H. Ong, C. Kang, J. H. Lim, J. Liu, D. Racoceanu, and K. Foong, "Cell clumping quantification and automatic area classification in peripheral blood smear images," *Proc. 2009 2nd Int. Conf. Biomed. Eng. Informatics, BMEI 2009*, no. c, pp. 2–6, 2009.

[23] R. W. R. Wang and B. McCane, "Red Blood Cell Classification through Depth Map and Surface Feature," *2008 Int. Symp. Comput. Sci. Comput. Technol.*, vol. 2, pp. 339–342, 2008.

[24] R. Cai, Q. Wu, R. Zhang, L. Fan and C. Ruan, *Red Blood Cell Segmentation using Active Appearance Model,* IEEE 11th International Conference on Signal Processing, pp. 1641-1644, Vol. 3, 2012.

[25] S. Kareem, R.C.S. Morling and I. Kale, *A Novel Method to Count the Red Blood Cells in Thin Blood Films,* IEEE, pp. 1021-1024, 2011.

[26] S.S. Barpanda, *Use of Image Processing Techniques to Automatically Diagnose Sickle-Cell Anemia Present in Red Blood Cells Smear,* Thesis, India, 2013.

[27] M. Hamghalam and A. Ayatollahi, *Automatic Counting of Leukocytes in Giemsa-stained Image of Peripheral Blood Smear,* IEEE International Conference of Digital Image Processing, 2009.

[28] Chen T., Zhang Y., Wang C., Qu Z., Cai M., Wang F. and S-M. Tanveer, *Local Complex Phase Based Level Set and Its Applications to DIC Red Blood Cell Segmentation.* IEEE-ISBI, pp. 187-190, 2011.

[29] C. Chandra Mouli, P. Jyothi, K. Nagabhushan Raju, *Comparative Study of Supervised Classification Algorithms for WOSRAS.* International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering.

[30] Razali Tomari, Wan Nurshazwani Wan Zakaria, Muhammad Mahadi, Abdul Jamil, Faridah Mohd Nor, Nik Farhan Nik Fuad. *Computer Aided System for Red Blood Cell Classification in Blood Smear Image*. International Conference on Robot PRIDE 2013-2014 - Medical and Rehabilitation Robotics and Instrumentation, ConfPRIDE 2013-2014.