



# **HYPERSPECTRAL IMAGE PROCESSING FOR CHRONIC MYELOID LEUKEMIA DIAGNOSIS**

**A THESIS** *Submitted in partial fulfillment of the*

*requirements for the award of the degree*

*of*

**DOCTOR OF PHILOSOPHY**

*by*

**AMRIT PANDA**



**DEPARTMENT OF ELECTRICAL ENGINEERING  
INDIAN INSTITUTE OF TECHNOLOGY  
INDORE**

**August 2024**



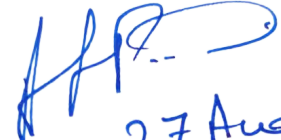


## INDIAN INSTITUTE OF TECHNOLOGY INDORE

### CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **“HYPER SPECTRAL IMAGE PROCESSING FOR CHRONIC MYELOID LEUKEMIA DIAGNOSIS”** in the partial fulfillment of the requirements for the award of the degree of DOCTOR OF PHILOSOPHY and submitted in the DEPARTMENT OF ELECTRICAL ENGINEERING, Indian Institute of Technology Indore, is an authentic record of my own work carried out during the time period from July 2016 to July 2023 under the supervision of Prof. Ram Bilas Pachori, Indian Institute of Technology Indore, India and Dr. Neeta Devi Sinnappah-Kang, Betty Cowan Research and Innovation Centre, Christian Medical College and Hospital, Ludhiana, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

  
27 Aug 24

Signature of the student with date  
(AMRIT PANDA)

-----

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.



Signature of Thesis Supervisor with date  
(Prof. RAM BILAS PACHORI)

-----

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

  
4/9/2024

Signature of Thesis Supervisor with date  
(Dr. NEETA DEVI SINNAPPAH-KANG)

-----

Amrit Panda has successfully given his Ph.D. Oral Examination held on **27.08.2023**



Signature of Thesis Supervisor

Date: **04.09.2024**



Convener, DPGC

Date: 04/09/2024

Signature of PSPC Member:1

Date:

Signature of PSPC Member:2

Date:

Signature of External Examiner

Date:

-----



## ACKNOWLEDGEMENTS

I am immensely grateful to my supervisor Prof. Ram Bilas Pachori and Dr Neeta Devi Sinnappah-Kang for their sustained inspiration, invaluable guidance, and unfailing support throughout my research work. It was my pleasure to work under the supervision and tutelage of such eminent persons who are established authorities in their respective fields.

I would like to express my sincere gratitude to Dr. Naveen Kakkar and Dr. M. Joseph John for their sagacious inputs pertaining to the biomedical aspects of the blood sample analysis and helping me develop a basic understanding of pathology. I am also grateful towards my PSPC committee members Dr. Aruna Tiwari and Dr. Vivek Kanhangad for their interesting discussions and suggestions during my research. I would like to thank my friends Dr. Vipin Gupta, Dr. Abhijit Bhattacharyya and Mr Kritiprasanna Das who helped me in various ways during my thesis work.

Above all, I am indebted to my late father, my mother, my wife, my son, my daughter and my sister for their unconditional love and support. I am specially grateful to my mother, without whose invaluable blessings and inspiration, I would not have been able to accomplish this significant milestone of my life. I am also thankful to my wife for her unflinching and selfless support and encouragement through out my research pursuit.

Last but not the least, I thank Lord Jagannath for having enlightened my mind to carry out this research work.

AMRIT PANDA





*Dedicated*

*to*

*My teachers and my family*



## ABSTRACT

Cancer is a broad term for the uncontrolled growth of cells. Chronic myeloid leukemia (CML) is a common type of cancer that affects myeloid cells which form blood in the body. In a CML patient, myeloid cells undergo a genetic mutation called translocation in which, a segment of the chromosome 9 detaches itself from its original sequence and attaches with a segment of chromosome 22 to form Philadelphia chromosome. It contains two genes fused together to form a composite mutated gene called BCR-ABL gene. The BCR-ABL gene causes CML by producing tyrosine kinase enzyme which triggers an abnormal growth in the number of white blood cells and reduction in the number of red blood cells in the blood.

Two methods were proposed based on the inherent spectral characteristics of the hyperspectral cube. The first method is the windowed Spectral angle mapping (SAM) method. When SAM was implemented over a group of test pixels in reduced bands, it is called windowed SAM method. In this method the classification was done based on the spectral angle between the training and the test pixels. The second method proposed was the 3- Dimensional Spectral Gradient Mapping (3-D SGM) method. The 3-D SGM method was proposed to overcome limitations in windowed SAM method by exploiting the inter-pixel and inter-band information content of the hyperspectral cube after carrying out necessary reduction in dimensions. This method not only exploits the spectral information but also uses the inter-pixel information to classify CML neutrophils. The classification carried out using 3-D SGM method was compared with windowed SAM method and analysed. The specificity measure for the algorithms showed that 3-D SGM (97.7%) was superior to windowed SAM (72.7%). It was also found to be superior to the Mahalanobis distance method (89.1%) in ruling out the presence of disease. The accuracy shown by the 3-D SGM (84.2%) was found to be fairly higher than the other methods. Also the high value of positive likelihood ratio indicated that the 3-D SGM was superior in diagnosing the presence of the disease (i.e., positive test for CML) versus windowed SAM method. However 3-D SGM had a very low sensitivity (38.5%).

The use of statistical distances like Euclidean distance and Mahalanobis distance computed in hyperspectral space has also been explored for classification of CML neutrophils. Both the methods have been compared to draw analysis of the advantages and disadvantages. The method employing Euclidean distance for classification, was superior when it came to the sensitivity (81.8%) measure as compared to all other methods. The Mahalanobis distance method was found to be better as compared to the Euclidean distance method for specificity (89.1%) and accuracy (82.5%) measures.

Two different methods based on the Probability distribution function (PDF) were proposed to improve the sensitivity, specificity, and accuracy values while classifying CML affected neutrophils from hyperspectral images. The first method is a combination of two methods based on PDF. It proposes a composite metric consisting of two parameters, i.e. the Mahalanobis distance and the Frequency domain normal PDF matching (FDNPM) parameter. This merged method exploits the benefits of both the Mahalanobis distance method which is computed in the spatial domain and the FDNPM method which uses the information content in the frequency domain. The merged method using the composite metric reduces the likelihood of missing diseased cases with a lower LR(-) value (0.4%) and improves the sensitivity (63.6%) without compromising the high specificity and accuracy values in the Mahalanobis distance method. In fact it further improves the specificity and accuracy values to 93.5% and 87.7% respectively. The second method is the Kurtosis compensated Euclidean distance (KCED) method which incorporates the effect of harmonics or tailedness of the distribution in the classification procedure. KCED significantly increases the specificity, accuracy and LR(+) values and reduces the LR(-) value, while maintaining specificity values as obtained in the Euclidean distance method. The KCED method demonstrates the highest sensitivity (81.8%), accuracy (91.2%) and LR(-) value (0.2). In addition, KCED method has the second best specificity (93.5%) and LR(+) value (12.6). Therefore, overall KCED was found to have demonstrated superior classification capability as compared to all other methods.



# Contents

<b>ABSTRACT</b>	<b>i</b>
<b>LIST OF FIGURES</b>	<b>ix</b>
<b>LIST OF TABLES</b>	<b>xiv</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xvi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 CML . . . . .	1
1.2 Hyperspectral Imaging . . . . .	4
1.2.1 Hyperspectral Image Sensors . . . . .	7
1.3 Blood Samples and Consumables . . . . .	8
1.3.1 Leishman Stain . . . . .	8
1.3.2 SpectraView Hyperspectral Imaging System . . . . .	11
1.4 Hyperspectral Image Processing . . . . .	13
1.4.1 Image Pre-processing . . . . .	13
1.4.2 Change in Data Structure and Dimension Reduction using Principal Component Transform . . . . .	14
1.4.3 Training Pixels . . . . .	24
1.4.4 Classification Techniques . . . . .	26
1.4.5 Evaluation Parameters . . . . .	27
1.5 Motivation . . . . .	28
1.6 Objectives . . . . .	29

1.7	Contributions . . . . .	30
1.8	Organization of the Thesis . . . . .	32
<b>2</b>	<b>Spectral Angle Mapping (SAM) and 3-D Spectral Gradient Mapping Method (3-D SGM)</b>	<b>35</b>
2.1	Introduction . . . . .	35
2.2	Spectral Angle Mapping (SAM) . . . . .	36
2.2.1	Implementation of SAM: Windowed SAM . . . . .	37
2.2.2	Shortcomings in SAM . . . . .	37
2.3	3-Dimensional Spectral Gradient Mapping (3-D SGM) . . . . .	39
2.3.1	Gradient Planes . . . . .	39
2.3.2	3-D Spectral Gradient Cube and 4-D Pixel Value of the Spectral Gradient Cube . . . . .	41
2.3.3	Spectral Gradient Cube of Training Pixels . . . . .	41
2.3.4	Implementation of 3-D SGM . . . . .	42
2.4	Results . . . . .	44
2.4.1	Sensitivity, Specificity and Accuracy . . . . .	44
2.4.2	Likelihood Ratios . . . . .	46
2.5	Discussion . . . . .	47
<b>3</b>	<b>Statistical Distances</b>	<b>50</b>
3.1	Introduction . . . . .	50
3.1.1	Classification using Euclidean Distance . . . . .	50
3.1.2	Limitations of Euclidean Distance for Classification . . . . .	52
3.1.3	Classification using Mahalanobis Distance . . . . .	56
3.1.4	Implementation of Mahalanobis Distance Method for Classification . . . . .	56
3.2	Results . . . . .	58
3.3	Discussion . . . . .	61
3.3.1	Comparison with Windowed SAM and 3-D SGM Methods . . . . .	62

3.3.2	Comparison between Euclidean and Mahalanobis Distance Methods . . . . .	62
<b>4</b>	<b>Classification of CML Neutrophils using Frequency Domain Normal Probability Distribution Function Merged with Mahalanobis Distance Method</b>	<b>65</b>
4.1	Introduction . . . . .	65
4.2	Use of Mahalanobis Distance Method in Classifying Multi-dimensional Images and its Limitations . . . . .	66
4.3	Probability distribution function (PDF) . . . . .	66
4.4	FDNPM . . . . .	67
4.5	Merged Method . . . . .	67
4.6	Implementation of Merged Method . . . . .	69
4.6.1	PDF of training pixels . . . . .	69
4.6.2	PDF of Test Pixels . . . . .	71
4.6.3	Matching frequency domain PDF of the training and test pixels	73
4.6.4	<b>Merging of Mahalanobis and frequency domain PDF matching methods</b> . . . . .	<b>73</b>
4.7	Results . . . . .	75
4.7.1	Sensitivity, specificity and accuracy . . . . .	75
4.7.2	Likelihood ratios . . . . .	75
4.8	Discussion . . . . .	76
<b>5</b>	<b>Classification of CML Neutrophils using Kurtosis Compensated Euclidean Distance</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.1.1	Classification using KCED . . . . .	80
5.1.2	Scaling of Euclidean distance using Kurtosis . . . . .	84
5.1.3	Uniformity of parameters . . . . .	88
5.1.4	Sensitivity, specificity and accuracy . . . . .	88



5.1.5	Likelihood ratios . . . . .	89
5.2	Discussion . . . . .	90
<b>6</b>	<b>Discussion</b>	<b>92</b>
<b>7</b>	<b>Conclusion</b>	<b>99</b>
7.1	Other Medical Applications of Hyperspectral Image Processing . . . .	100
7.1.1	Hyperspectral Imaging for Analysing Biomedical Tissues . . .	101
7.1.2	Hyperspectral Imaging of Blood Samples . . . . .	101
7.2	Challenges in Use of Hyperspectral Images for Medical Diagnosis . . .	102
7.3	Future Scope of Work . . . . .	103



# List of Figures

1.1	A depiction of the hyperspectral image cube showing the images associated with various wavelength bands. . . . .	6
1.2	A block diagram of SpectraView hyperspectral imaging system showing its various components and modules. . . . .	9
1.3	Synthesis of methylene blue using dimethyl-4-phenylenediamine and sodium thiosulphate. . . . .	10
1.4	Representation of the chemical structure of Eosin Y (a) and Eosin B (b). . . . .	11
1.5	A block diagram showing the working of sagnac interferometer showing the segregation of spectrum in terms of wavelength based on phase difference induced by path difference. . . . .	12
1.6	An RGB image of a region of interest of the first CML blood sample slide with four neutrophils labeled as 'N' (Magnification: 400x). . . .	15
1.7	A flow chart showing the steps involved in carrying out PCT. . . . .	17
1.8	The graphical depiction of reduction in variance values of the hyperspectral image cube (the RGB image is shown in Fig. 1.2.) corresponding to higher PCT bands. . . . .	21
1.9	Spectral signature of first hyperspectral cube in the first seven bands after applying PCT. . . . .	22
1.10	The images in the first six wavelength bands of the hyperspectral image cube in descending order given from (a) to (f). . . . .	23

1.11	Spectral signature of (a) the cytoplasm and (b) the nucleus part of CML neutrophils in the seven maximum information bands after applying PCT. . . . .	25
2.1	Flowchart showing the implementation of Windowed SAM method for classification of hyperspectral image. . . . .	38
2.2	Gradient planes perpendicular to the X, Y and Z axis giving each pixel as a gradient vector with three components corresponding to the three planes. . . . .	40
2.3	Flowchart showing the implementation of 3-D SGM method for classification of hyperspectral image. . . . .	43
2.4	Detection of CML neutrophil in the sample <i>cmlcube3s_2.2</i> . (a) The RGB depiction of the CML cube; (b) the composite spectral gradient difference cube of cytoplasm and (c) nucleus; (d) the binary image showing the detected pixels for neutrophils. . . . .	45
3.1	Flowchart showing the implementation of Euclidean distance based method for classification of hyperspectral image. . . . .	53
3.2	The Euclidean distance of a sample pixel from the centroid of the given balanced or symmetric distribution in an n-D space. . . . .	54
3.3	The Euclidean distance of a sample pixel from the centroid of the given unbalanced or non-symmetric distribution in a n-D space. . . .	55
3.4	Flowchart showing the implementation of Mahalanobis distance based method for classification of hyperspectral image. . . . .	57

3.5	Detection of CML neutrophil in two samples. (a) The RGB depiction of the third hyperspectral sub-cube of the first hyperspectral cube sample; (b) The RGB depiction of the third hyperspectral sub-cube of the second hyperspectral cube sample; (c) The detection of CML neutrophils in the third hyperspectral sub-cube of the first hyperspectral cube sample using Euclidean distance method; (d) The detection of CML neutrophils in the third hyperspectral sub-cube of the second hyperspectral cube sample using Euclidean distance method; (e) The detection of CML neutrophils in the third hyperspectral sub-cube of the first hyperspectral cube sample using Mahalanobis distance method; (f) The detection of CML neutrophils in the third hyperspectral sub-cube of the second hyperspectral cube sample. . . . .	59
4.1	Flowchart showing the implementation of FDNPM method for classification of hyperspectral image. . . . .	68
4.2	Flowchart showing the implementation of Merged method comprising of the Mahalanobis distance method and the FDNPM method for classification of hyperspectral image. . . . .	70
4.3	The depiction of (a) RGB image of the third hyperspectral sub-cube obtained from the first hyperspectral cube and (b) the merged method (using Mahalanobis distance and FDNPM methods). . . . .	74
5.1	The sample $x$ is part of the heavy-tailed PDF in (a) but not a part of the light-tailed PDF in (b). The use of same Euclidean distance for PDF with different Kurtosis may result in incorrect classification results [ $d$ = distance of $x$ from centre of the normal PDF] . . . . .	85
5.2	Flowchart showing the implementation of KCED method for classification of hyperspectral image by applying kurtosis compensation on Euclidean distance. . . . .	86

5.3	The PDF across seven PCT bands [(a) to (g)] of one set (as a representation) of training pixels for cytoplasm (dotted line) and nucleus (solid line) of CML neutrophils. . . . .	87
5.4	The depiction of (a) RGB image of the third hyperspectral sub-cube obtained from the first hyperspectral cube, (b) the merged method (using Mahalanobis distance and FDNPM methods) and (c) the detection of CML neutrophils using KCED. . . . .	88



# List of Tables

2.1	Sensitivity, specificity and accuracy of Windowed SAM versus 3-D SGM algorithms. . . . .	46
2.2	Likelihood ratios of Windowed SAM and 3-D SGM algorithms. . . . .	47
3.1	The classification of hyperspectral data of the CML blood sample using Euclidean and Mahalanobis distances. . . . .	60
3.2	Likelihood ratios of classification using Euclidean and Mahalanobis distance methods. . . . .	60
4.1	Sensitivity, specificity and accuracy of the KCED and the merged (Mahalanobis distance + FDNPM) methods on healthy versus CML neutrophils. . . . .	75
4.2	Comparison of likelihood ratios of the Mahalanobis distance method and the merged (Mahalanobis distance + FDNPM) methods. . . . .	76
5.1	Sensitivity, specificity and accuracy of the KCED method on healthy versus CML neutrophils. . . . .	89
5.2	Likelihood ratios of the KCED method. . . . .	90
6.1	Comparison of Sensitivity, Specificity and Accuracy Values. . . . .	94
6.2	Comparison of Likelihood ratios. . . . .	94





# List of Abbreviations and Symbols

3-D SGM:	3-Dimensional spectral gradient mapping
ALL:	Acute Lymphoblastic Leukemia
BC:	Blast Crisis
CBC:	Complete blood cell count
CCD:	Charged coupled device
CLL:	Chronic lymphocytic leukemia
CML:	Chronic myeloid leukemia
CT:	Computer tomography
FDNPM:	Frequency Domain normal probability distribution function method
FN:	False Negative
FP:	False Positive
FISH:	Fluorescent in situ hybridization
KCED:	Kurtosis compensated Euclidean distance
LR:	Likelihood Ratio
PCA:	Principal component analysis
PCT:	Principal component transform
PDF:	Probability distribution function
SAM:	Spectral angle mapping
TKI:	Tyrosine Kinase Inhibitors
TN:	True Negative
TP:	True Positive

WSAM: Windowed spectral angle mapping  
OPD: Optical path difference  
RGB: Red green blue  
RMS: Root Mean Square  
RoI: Region of interest  
WBC: White blood cells



# Chapter 1

## Introduction

Cancer is a broad term for the uncontrolled growth of cells. CML is an uncommon type of blood cancer that originates inside the bone marrow and affects the myelogenous cells. It is a slow progressing disease mostly found in older adults, although its occurrence among children is not rare.

Timely detection is of great importance in treating CML. The study of the blood sample through naked human eye is not possible due to limited spectral and spatial resolution capability of human vision. To overcome these challenges it is necessary to capture the object with high spatial and spectral resolution. After obtaining the images, it is important to process them correctly to extract useful information. One such technique is hyperspectral imaging. Hyperspectral images can be considered as multiple stacks of images of the same object captured by hyperspectral sensors within a range of near contiguous wavelengths with an inter-band gap of 2 - 10 nm [9].

### 1.1 CML

Over 62 years ago the presence of a particular chromosomal abnormality linked to CML, was discovered [1,2]. In a CML patient, myeloid cells undergo a genetic mutation called translocation in which, a segment of the chromosome 9 detaches itself

from its original sequence and attaches with a segment of chromosome 22. It occurs in the haemopoietic stem cell (abbreviated as HSC) compartment in the medulla of the bone (bone marrow). This unique abnormality is a t(9;22) reciprocal chromosomal translocation and it was subsequently named as the Philadelphia chromosome [1-3]. It contains two genes fused together to form a composite mutated gene called BCR-ABL1 gene. The BCR-ABL1 gene characterises the cell biology of CML affected blood cells and is responsible for the cellular features of the leukaemia i.e., inhibition of apoptosis, enhanced cell growth, growth factor independence, impaired genomic surveillance, altered cell adhesion and differentiation [4]. The BCR-ABL1 gene produces Tyrosine Kinase enzyme. Tyrosine Kinase enzyme reprograms the White Blood Cells (WBC) causing myeloid hyperplasia which triggers an abnormal growth in the number of WBCs and reduction in the number of Red Blood Cells (RBC) in the blood. RBC are the carriers of oxygen to different parts of the body where they are used to burn the calories and generate energy for various biological and physiological functions. The reduction of RBC results in reduced availability of oxygen to various parts of the body resulting in lack of energy. Therefore, the prominent symptoms of CML are weakness and fatigue, fever, cachexia, excessive sweating, bleeding and an enlarged spleen.

The diagnosis of about 85% of CML patients is generally when the disease is in its chronic phase. Being asymptomatic, about half of these patients get diagnosed when their blood samples are analyzed for an unrelated medical examination [5]. Thus CML can go unnoticed and undiagnosed for years. In the chronic phase, CML is often presumed from the result of a complete blood count, where the white blood cell (WBC) count is elevated. In order to confirm the diagnosis, the characteristic Philadelphia chromosome is detected by routine cytogenetics and fluorescent in situ hybridization as well as quantitative real-time PCR for the BCR-ABL1 fusion gene. There are several studies that have shown case reports of patients who had prolonged “preleukemic” phases before developing CML (to name a few: [6-8]). Even though the presence of the Philadelphia chromosome was confirmed in these patients, they

had normal to mildly elevated WBC counts. They were classified as “preclinical”, “preleukemic” or “smoldering” CML patients.

It is not known what initiates the genetic mutation that sets off the development of CML. It is neither inherited nor preventable. Once the disease progresses to an advanced stage there is possibility of it assuming an aggressive phenotype. This can lead to a condition similar to acute leukemia called Blast Crisis (BC) and can become fatal. Without therapeutic intervention, within three to five years after the chronic phase, the disease will progress into BC resembling an acute leukemia. Early detection coupled with targeted treatment against this abnormal protein is a viable way to treat this disease.

The diagnostic methods for CML include traditional methods of diagnosis like Complete Blood Count (CBC) test, molecular testing using cytogenetics and biopsy and bone marrow aspiration. The treatment for CML primarily involves administration of Tyrosine Kinase Inhibitors (TKI) to achieve deeper cytogenetic/ molecular responses. Over a longer period of time this aids in achieving better outcomes and mitigating the rate of progression of CML [9-10].

The visual microscopic assessment of peripheral blood smears are often time consuming and limited (i.e., a fixed number of cells are examined per slide), and is subjective in nature. Yet this method is more economical (especially in developing countries like India) versus the molecular-based techniques stated above. Many studies have implemented the use of computer-aided methods to analyze RGB images of blood smears [11]. Shafique and Tehsin used online database images of acute lymphoblastic leukemia (ALL) for their work [9,12,13]. They used 60% of their data for training and 40% for testing [9]. They achieved an accuracy of 96.06% in classifying ALL using deep convolutional neural network. Mohapatra et al. analyzed morphological and textural characteristics from RGB images of peripheral blood and bone marrow using image processing. They developed a quantitative microscopic method to differentiate lymphoblasts from healthy lymphocytes [14], and applied machine vision and machine learning techniques to automate the detection of lym-

phoblasts. Their investigation showed that they could achieve a 99% accuracy by using the multi-classifier approach [14-15].

The presence of leukemia cells is characterized by an unusual increase in the number of white blood cells. Neutrophils are the most abundant (60 – 70%) cell type in blood circulation. It performs an important role in the innate immunity by providing the first line of defense [16,17]. Therefore, in the present work classification of CML affected neutrophils has been considered to diagnose CML patients from their blood samples.

## 1.2 Hyperspectral Imaging

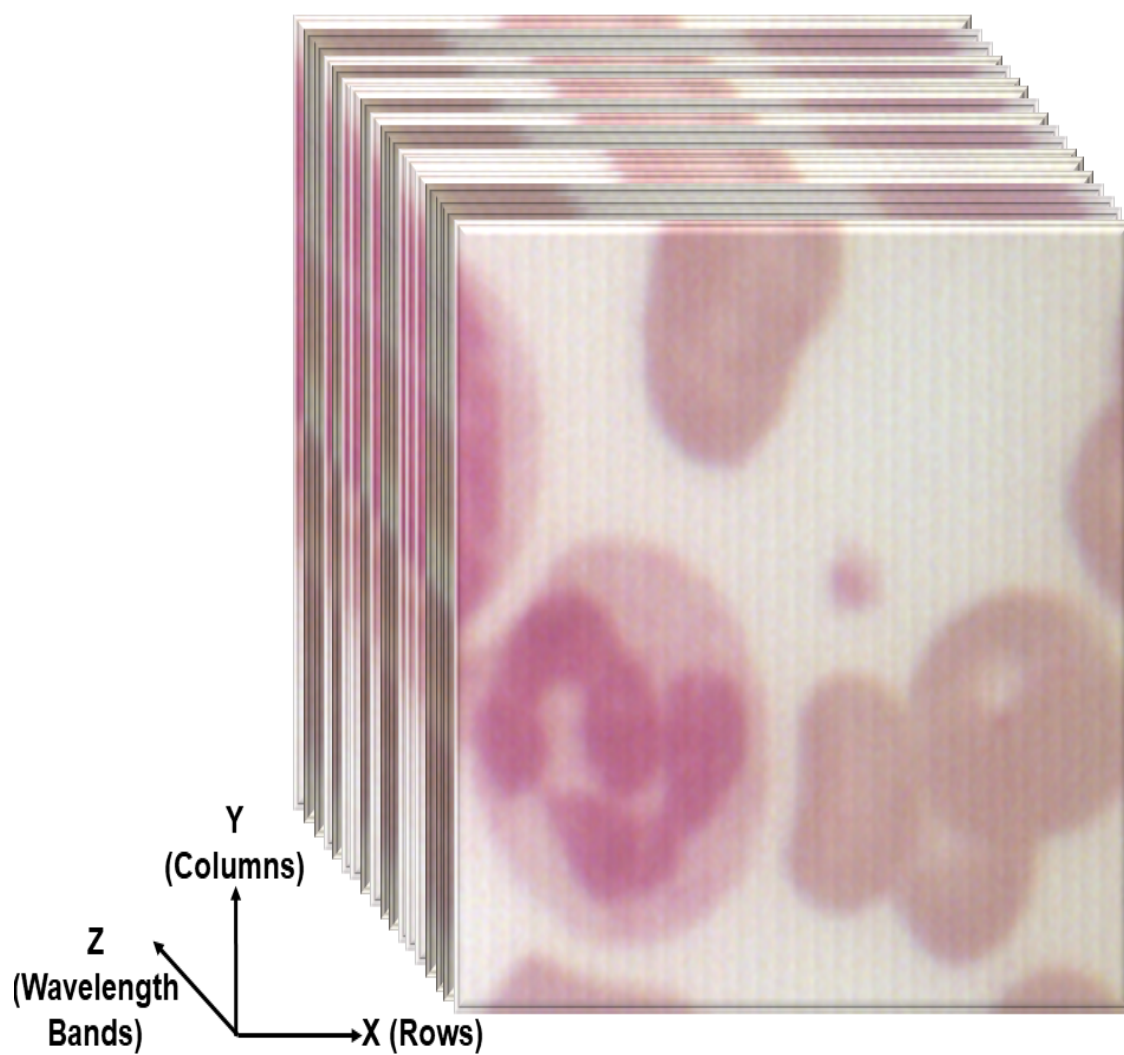
Study of the blood sample through naked human eye is not possible due to limited spectral and spatial resolution of human vision. In addition to it the human vision is over a limited range of wavelength bands. However, critical information for classification of a sample may actually be in wavelength bands beyond the human visible spectra. To overcome these challenges it is necessary to capture the object over a larger spectral range with high spatial and spectral resolution. After obtaining the images, it is important to process them correctly to extract useful information. With advances made in imaging technology, it is possible to sense objects at high spatial as well as spectral resolutions and also wide spectral ranges. One such technique is hyperspectral imaging. Hyperspectral images can be considered as multiple stacks of images of the same object captured by hyperspectral sensors within a range of near contiguous wavelengths with an inter-band gap of 2-10 nm [18-20]. The illustrative hyperspectral image cube with stacks of images corresponding to various wavelengths is shown in Fig. 1.1.

Historically, multispectral and hyperspectral image processing were extensively researched in the field of remote sensing [19,21-32]. The use of hyperspectral images for classifying cells and tissues in the field of biomedicine has gained popularity as it combines imaging with spectroscopy [24]. Hyperspectral image are multidimensional



images captured over multiple wavelength bands with very narrow band gaps. The image slices captured at individual wavelength bands are independent of each other. Therefore, a hyperspectral image pixel can be considered as a linear-scaled mixture of orthogonal intensities captured at each wavelength band [19,21-22,33]. Biomedical applications of hyperspectral imaging include analysis of dried blood spots [34], lymphoproliferative disorders [35], morphometric characterization of B-Chronic lymphocytic leukemia (CLL) [35], spatial features of colon biopsy [36], spectral-spatial features to differentiate lymphoblasts from lymphocytes [37], and white blood cells [38]. The images generated using this technology are large in size because of the combined high spectral and spatial resolutions [32]. Thus, processing such huge data requires some adaptation and/or merger of algorithms. Thus there is tremendous potential for non-invasive diagnosis of diseases with the entry of hyperspectral imaging into the field of medical applications [39].

One of the initial applications of hyperspectral sensors was in the field of remote sensing and military target detection [21-23,26,32,37]. The motivation for using hyperspectral imaging for such applications was the inherent low spatial resolution of the images due to the huge distances between the object and the sensor. For example a camera placed in a satellite orbiting in the geo-stationary orbit at an approximate distance of 36000 km above the surface of the earth would produce very low spatial resolution images [19,25,30,40]. It is possible to extract useful information from such low spatial resolution images by increasing the spectral resolution [15,19,21]. This was achieved by using multispectral cameras which captures the images in multiple wavelength bands. Hyperspectral imaging goes a step forward by reducing the band gap between the wavelengths to as less as 10 nm. Such low band gaps result in the image being captured over near contiguous bands of wavelengths. Hyperspectral imaging has been extensively researched and used in remote sensing for land use mapping, mapping underground mineral deposits, analysing crop cover and forest cover. It finds special applications in the field of military target detection under camouflaged conditions. Hyperspectral image processing techniques have also been



**Figure 1.1** A depiction of the hyperspectral image cube showing the images associated with various wavelength bands.

used for anomaly detection and material analysis [41,42].

Of late hyperspectral imaging has made its foray in the field of bio-medical image processing. Hyperspectral images of bio-medical samples can be captured by mounting the hyperspectral sensor over a microscope. Unlike remotely sensed hyperspectral data, the hyperspectral images of biomedical samples are of high spectral as well as spatial resolution. Such high spatial and spectral resolution hyperspectral image cubes are information rich and form large data sets. Analysis of such heavy and complex data sets to extract useful information requires robust algorithms. Several studies have analyzed hyperspectral images of various blood cells using machine learning approaches [43]. Hyperspectral imaging has been used to analyze dried blood spots [34], lympho-proliferative disorders [35], morphometric characterization of B-Chronic lymphocytic leukemia (CLL) [36], spatial features of colon biopsy, spectral-spatial features to differentiate lymphoblasts from lymphocytes [37-38], and white blood cells, to name a few. The images generated using this technology is large in size because of the combined high spectral and spatial resolutions [32]. Thus, processing such huge data requires some adaptation and/or merger of algorithms.

### **1.2.1 Hyperspectral Image Sensors**

Hyperspectral image sensors are cameras that capture the light in various wavelengths to generate a spectrum of each object pixel. In the present study, the hyperspectral cubes were captured using the SpectraView hyperspectral imaging system. This imaging system consists of the spectra cube, Charge Coupled Device (CCD) sensor and an interferometer. The spectra cube splits the incident beam along two different paths, so as to induce an Optical Path Difference (OPD). At the output of the spectra cube the incident beam is combined together with the OPD. The OPD results in generation of image frames which are captured using the CCD camera. The image frames are used to build interferograms corresponding to each

image pixel. The spectrum of the image pixels are obtained by computing the fourier transform of the interferogram [19,22]. The block diagram of the SpectraView hyperspectral imaging system consisting of a sagnac interferometer for obtaining the spectra is shown in Fig. 1.2.

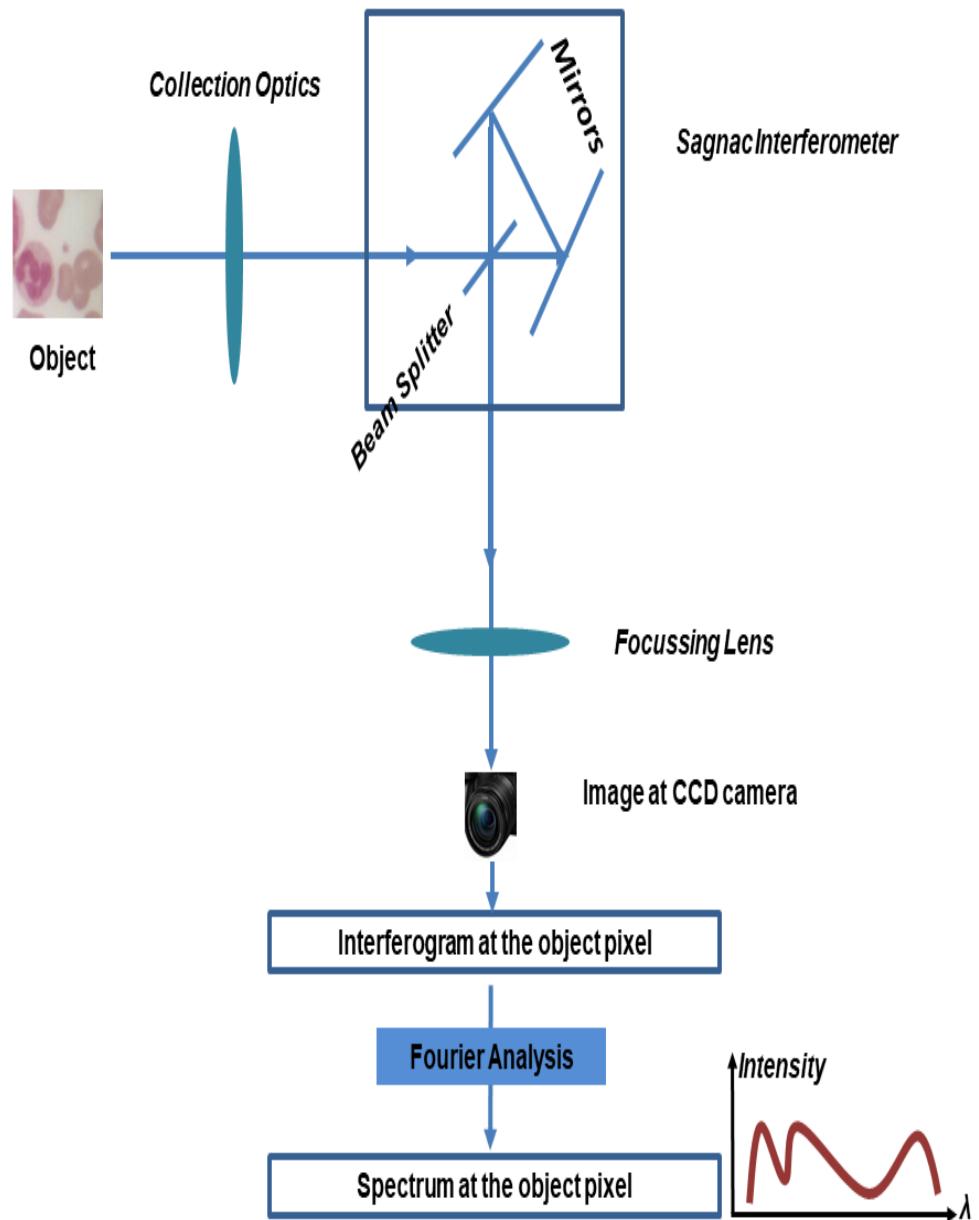
## **1.3 Blood Samples and Consumables**

### **1.3.1 Leishman Stain**

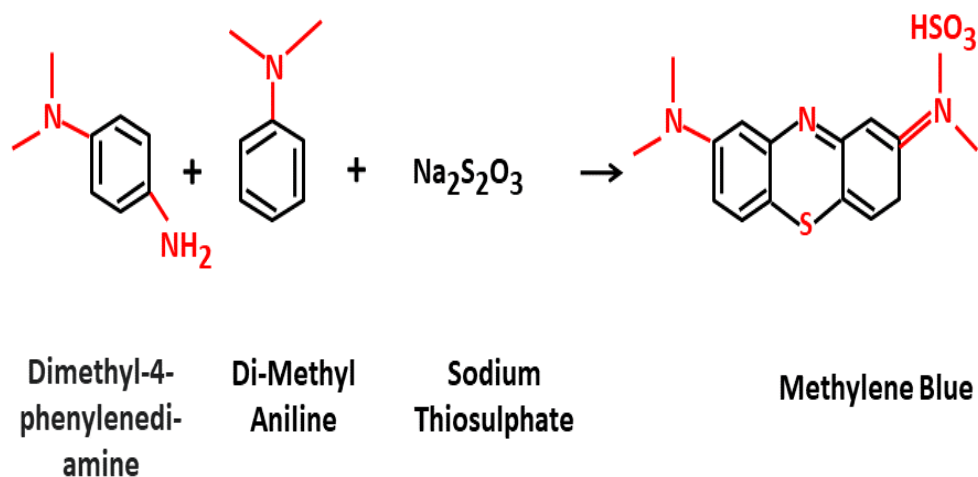
In the present study, a CML and a healthy blood smear sample were stained with Leishman stain (Merck, Germany). Leishman stain is based on methanolic mixture comprising of a basic dye, i.e., methylene blue and an acidic dye, i.e., eosin. Proper staining formed an important part of the entire procedure of preparation of blood smear slides. The slides which were prepared had non-overlapping cells and were stained with optimised amount of the Leishman stain, such that overstaining as well as understaining was avoided.

#### **Methylene Blue**

Methylene blue is also called Methylthioninium chloride. It has widespread uses which include use as a drug for medication and a dye for staining tissues and pathological blood samples. The ability of methylene blue to convert ferric iron present in haemoglobin into ferrous iron by a process of reduction has been exploited to treat methemoglobinemia. Methylene blue is a green powder which yields a blue solution in water and is prepared as a formal derivative of phenothiazine. The exact preparation procedure for methylene blue comprises of oxidation of dimethyl-4-phenylenediamine in presence of sodium thiosulphate as given at Fig. 1.3.



**Figure 1.2** A block diagram of SpectraView hyperspectral imaging system showing its various components and modules.

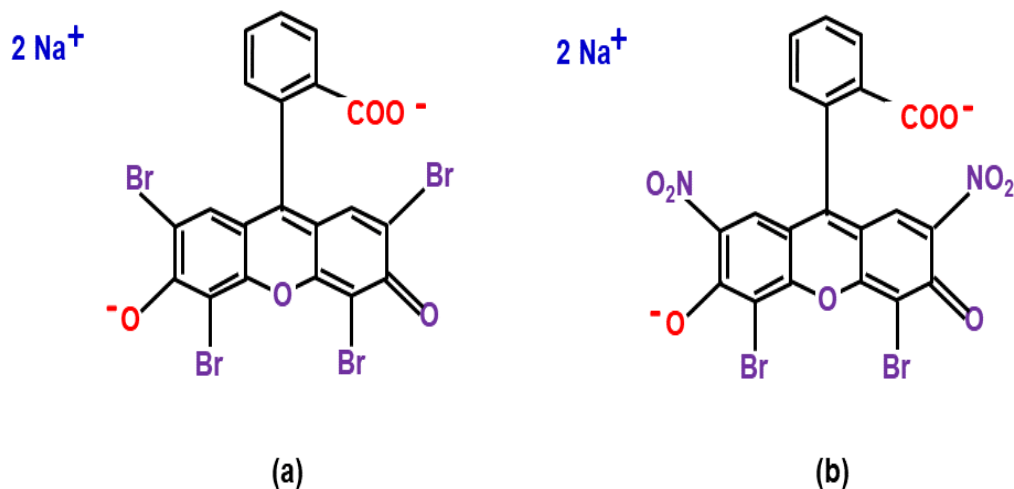


**Figure 1.3** Synthesis of methylene blue using dimethyl-4-phenylenediamine and sodium thiosulphate.

## Eosin

Eosin is a fluorescent acidic compound which binds to basic or alkaline compounds in order to form salts. The salts are generally dark red or pink. This is due to the action of bromine on eosin. The resultant salt has a property of staining proteins in the cytoplasm. There are two types of eosins, namely, Eosin B and Eosin Y. Eosin B is a di-bromo di-nitro derivative of fluorescein, whereas Eosin Y is a tetrabromo derivative of fluorescein. The chemical structure of Eosin Y and Eosin B is given at Fig. 1.4. For the purpose of histological staining Eosin Y is the most commonly used.

In Leishman stain, methylene blue binds to the acidic region of a cell and eosin binds to the basic region of the cell. Regions of interest (ROI) on the slides where cells did not overlap each other were used for capturing hyperspectral image cubes. The images were mostly captured in the middle region of the blood smear where density of cells were sufficient without overlap. The samples were part of a bigger project, that has been approved by the Christian Medical College Research and Ethics Committee. The samples were coded to protect the identity of the CML patient and the healthy volunteer.



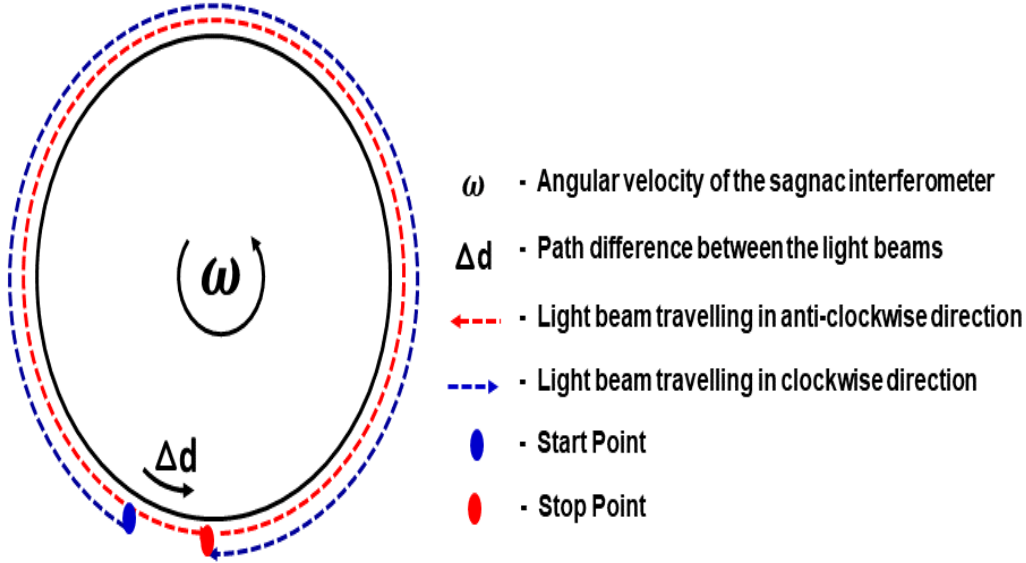
**Figure 1.4** Representation of the chemical structure of Eosin Y (a) and Eosin B (b).

### 1.3.2 SpectraView Hyperspectral Imaging System

The hyperspectral image cubes were captured from the blood samples (by NDSK<sup>1</sup>) at Betty Cowan Research and Innovation Centre using the SpectraView Hyperspectral Imaging System (Applied Spectral Imaging, Israel) [44-45]. The working of SpectraView Hyperspectral Imaging System is based on the principle of interferometry and Fourier transform. Specifically, it uses the Sagnac interferometer. Sagnac interferometer is a ring interferometer which splits the light into two different parts and makes them travel in two opposite directions along a ring. The light beams are made to reach a common detector co-located with the source. The path travelled by the light beams would be the same if the ring through which the light is made to travel is static. However, if the ring is made to rotate, then the paths travelled by both the beams would be different as shown in Fig. 1.5. This is because the relative velocity of the light beam increases when the beam is travelling opposite to the direction of rotation and decreases when the light is travelling along the direction of rotation of the ring. This results in phase difference when the beams of

---

<sup>1</sup>Neeta Devi Sinnappah-Kang



**Figure 1.5** A block diagram showing the working of sagnac interferometer showing the segregation of spetrum in terms of wavelength based on phase difference induced by path difference.

light reach the detector. The phase difference of different wavelengths of light is different. These light beams are made to merge together and undergo interference. The interfering beams form interfering patterns called fringes. The gap between the interfering fringes is decided by the angular velocity of rotation of the interferometer apparatus. The interfering patterns give the various wavelengths of light in spatial domain. The constituent wavelengths of light can be distinguished by carrying out a fourier transform of the interference patterns. The schematic diagram of the SpectraView Hyperspectral Imaging System is given in Fig. 1.5.

The Region of Interest (RoI) was identified on the blood sample slides, so as to ensure capture of maximum density of non-overlapping blood cells. The hyperspectral image cubes of these RoI were generated using the SpectraView imaging system. The hyperspectral image cubes were coded to ensure confidentiality with respect to the identity of the persons from whom the blood samples were collected. Thirteen hyperspectral image cubes were captured from the blood samples. These image cubes were spatially split to obtain 57 hyperspectral image sub-cubes. The healthy volunteers contributed to 32 image sub-cubes and the CML affected patients contributed to a total of 25 image sub-cubes. The hyperspectral cubes were



obtained across 61 wavelength bands for all images (400.16-1018.39 nm) except one which was captured over 49 wavelength bands (402.34-804.71 nm). After reduction in dimensions, the data from the visible range of wavelengths was found to have the maximum amount of information [44-45]. Therefore, the final outcome of the results were not affected by these differences in the wavelength bands [21,24,45].

## 1.4 Hyperspectral Image Processing

Hyperspectral image cubes are in the form of huge data sets which can be interpreted as image planes in multiple dimensions corresponding to each wavelength. Since the hyperspectral data sets are huge they require special techniques for processing. First the raw image cube needs to be normalised, the minimum noise level needs to be identified and removed. Since the hyperspectral image cube is a huge data set, it needs to be subjected to dimension reduction before applying classification algorithms.

### 1.4.1 Image Pre-processing

The intensity values in the hyperspectral cube were obtained based on the raw spectra captured. Since the distance and angle of the electromagnetic radiation from the object to the sensor was different for different pixels, the normalization of the image cube was carried out with respect to the cube maxima and then standardised to an 8 bit radiometric resolution. The maximum intensity of the pixel in any band, obtained by this method was 255. A CCD camera has some amount of dark current which manifests as a minimum intensity value in the image cube [21,24]. To ensure zero minima the image cube was further normalised by subjecting it to dark area subtraction. The intensity of the pixels after normalisation and dark area subtraction is given by equation (1.1).

$$I_{nd} = \left( \frac{I - I_{min}}{(I_{max})} \right) \times 255 \quad (1.1)$$

where,

$I_{nd}$  - Normalised pixel intensity obtained after dark area subtraction.

$I$  - Original pixel intensity in a given wavelength.

$I_{max}$  - Maximum pixel intensity in the entire image cube.

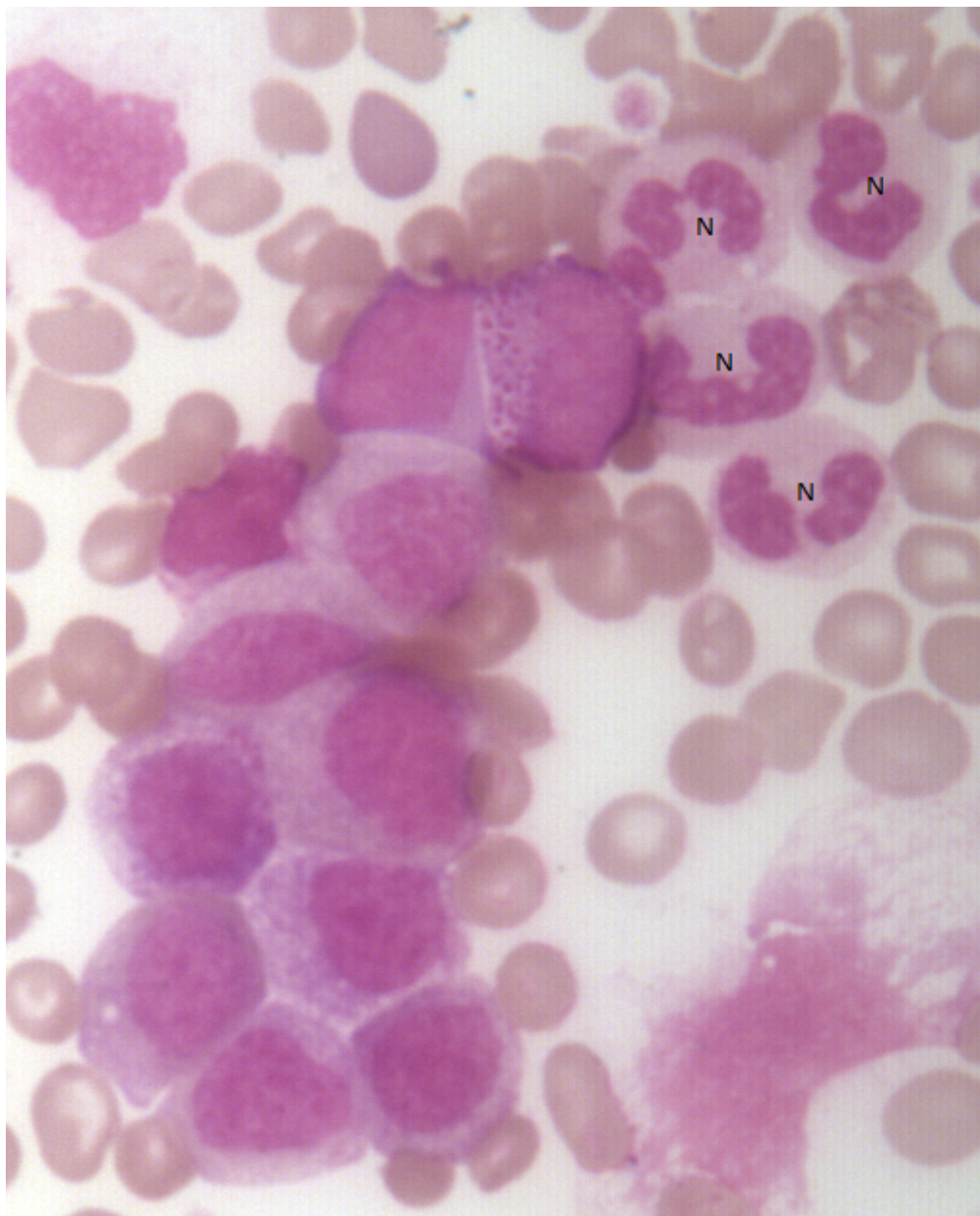
$I_{min}$  - Minimum pixel intensity in the entire image cube.

The resultant hyperspectral image cube consisted of multiple stacks of pre-processed images corresponding to each hyperspectral wavelength. The three band image in respect of the third hyperspectral image sub-cube of the first hyperspectral image cube obtained by combining the images from the wavelengths corresponding to the red , green and blue wavelengths is shown in Fig. 1.6. with the neutrophils marked as N .

### 1.4.2 Change in Data Structure and Dimension Reduction using Principal Component Transform

#### Reduction in Dimensions

Hyperspectral images are heavy datasets consisting of image slices in multiple wavelengths. Processing such heavy datasets is processor intensive and hence time consuming. In order to efficiently process such heavy datasets, it is essential that the dimensions be reduced without effectively compromising the information content. many types of data reduction techniques exist like minimum noise fraction transform, independent component analysis etc. In the present work, Principal Component Analysis (PCA) has been used to achieve effective dimension reduction. PCA is beased on Principal Component Transform (PCT).

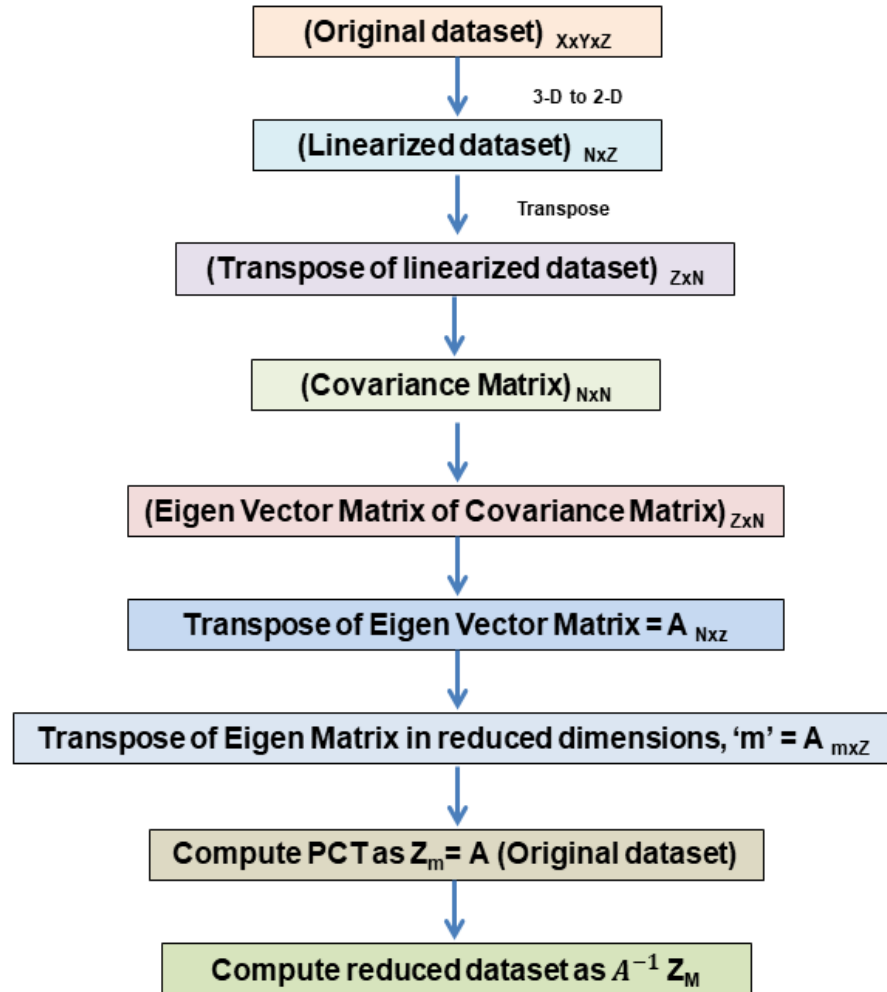


**Figure 1.6** An RGB image of a region of interest of the first CML blood sample slide with four neutrophils labeled as 'N' (Magnification: 400x).

## Principal Component Transform (PCT)

The format in which the data set was obtained by the sensor was in the form of a ‘.raw’ extension file. For ease of applying the proposed methods the pre-processed data set was further processed using MATLAB to generate a format with number of rows as first dimension, number of columns as second dimension and number of bands as third dimension. Processing such high spectral and spatial resolution data sets was processor intensive. Therefore it was necessary to decompose the hyperspectral dataset in a manner that the high information containing wavelengths can be segregated from the low information containing wavelengths. Various methods for decomposing the hyperspectral dataset were explored [46]. 7 high information containing bands out of 61 or 49 bands, as applicable were identified for each image using principal component analysis (PCA) [21,24-28,45] after carrying out Principal Component Transform (PCT). PCA is a statistical procedure which generates a set of uncorrelated orthogonal variables called the principal components. In the present work, the available data set was transformed to a linear scaled mixture of the principal components [19]. The principal components are essentially eigenvectors of the covariance matrix of the data set. Since the covariance matrix is symmetric, the principal components are orthogonal to each other. The principal component transformation resulted in the rotation of the vector axis system in the direction of the eigenvectors. This resulted in a high projection value of the data in the new vector coordinate system. The principal component corresponding to the highest eigenvalue is the first principal component and was found to display the highest variance and highest information content [19,44-46]. The flow chart demonstrating the use of PCT to achieve reduction in dimensions is given in Fig. 1.7.

In the present study, the  $x \times y \times 61$  cube data set was translated into an  $N \times 61$  data set spread across 61 dimensions, where  $N$  is the total number of vector pixels in the image and equals to  $x \times y$ . The translated linear hyperspectral image in all its 61 dimensions was depicted as the variable  $X_{61 \times N}$ . The  $N \times N$  covariance matrix was built up from the data set given at equation (1.2).



**Figure 1.7** A flow chart showing the steps involved in carrying out PCT.

$$X_{61 \times N} = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & x_{1N} \\ x_{21} & x_{22} & \cdot & \cdot & x_{2N} \\ \cdot & & & & \\ \cdot & & & & \\ x_{611} & x_{612} & \cdot & \cdot & x_{61N} \end{bmatrix} \quad (1.2)$$

where,

$x_{i,j}$ - The intensity value of the  $j^{th}$  pixel in the  $i^{th}$  band.

$N$ - The total number of pixels in the hyper spectral image and ranges from one to  $x \times y$ .

An  $N \times N$  covariance matrix was built up from the data set given at equation (1.2) and was denoted as  $\Sigma_X$ , as given in equation (1.3).

$$\Sigma_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \sigma_{2N} \\ \cdot & & & & \\ \cdot & & & & \\ \sigma_{N1} & \sigma_{N2} & \cdot & \cdot & \sigma_{NN} \end{bmatrix} \quad (1.3)$$

thus,

$$\sigma(i, j) = E(\Sigma_{i=1}^N \Sigma_{j=1}^N (x_i - \bar{x}_i)(x_j - \bar{x}_j)^T)$$

where,

$\sigma(i, j)$ - The covariance value computed between vector pixels  $x_i$  and  $x_j$ , for all  $i$  and  $j$ .

$x_i$ - The  $i^{th}$  pixel out of the total  $N$  pixels of the hyperspectral image.

$x_j$ - The  $j^{th}$  pixel out of the total  $N$  pixels of the hyperspectral image.

Next an  $N \times N$  eigenvalue matrix of the covariance matrix, denoted by  $\lambda$ , was derived for the eigenvector matrix,  $A$ . The eigenvalue matrix,  $\lambda$  was a diagonal matrix, with the diagonal formed by the eigenvalues of the covariance matrix  $\Sigma_X$

[37,38]. Eigenvector matrix obtained was an orthonormal matrix composed of the 61 dimensional or 49 dimensional eigenvectors (as applicable) of the covariance matrix  $\Sigma_X$  and was denoted as:

$$A_{61 \times N} = \begin{bmatrix} a_{11} & a_{12} & . & . & a_{1N} \\ a_{21} & a_{22} & . & . & a_{2N} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{611} & a_{612} & . & . & a_{61N} \end{bmatrix} \quad (1.4)$$

where,

$a_{i,j}$ - The  $j^{th}$  eigen vector in the  $i^{th}$  band.

The eigenvector matrix can be written with column depicting bands as follows:

$$A_{N \times 61}^T = \begin{bmatrix} a_{11} & a_{21} & . & . & a_{611} \\ a_{12} & a_{22} & . & . & a_{612} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{1N} & a_{2N} & . & . & a_{61N} \end{bmatrix} \quad (1.5)$$

The diagonal spread of the data set is given by the covariance matrix. The eigenvectors of the covariance transformation matrix represent the unit vectors along the direction of the spread. The eigenvalues define the magnitude by which the eigenvectors need to be scaled. When the eigenvector matrix is rearranged in decreasing order of the corresponding eigenvalues, it forms the PCT matrix. A vector can be decomposed into its constituent vectors by applying PCT. The decomposed constituents are in decreasing order of their variances/ information content. Therefore, the PCT of any vector would give a linear scaled combination of the constituent principal components. The scaling of the principal components is done by their respective eigenvalues [22]. The same is given at equation (1.6).

$$A^T(X) = \lambda X \quad (1.6)$$

PCT was used to translate the original data set to a sum of principal components which can be depicted as linear weighted sum of the original data set. The weights were the eigenvalues of the covariance matrix of the original data set. The eigenvalues being in decreasing order, the variances of the principal components were also in decreasing order. Since the variance is an indicator of the information content of the principal component, the information content of the data set can also be assumed to be in decreasing order of their principal components [21,28,44-45]. Therefore, the first image had maximum information content, where the inter pixel variance was maximum. This reduced for the higher PCT wavelength bands. The graphical representation of the variance or information content of the images corresponding to  $61 \times$  PCT bands is graphically represented in Fig. 1.8.

If the PCT of any vector,  $X$  is  $Z$ , i.e.,  $Z = AX$ , then the PCT in the first  $m$  bands can be obtained by taking the first  $m$  bands or rows of the eigenvalue matrix  $\lambda$ , i.e.,  $\lambda_m$ . The resultant PCT of the vector,  $X$ , is given by  $Z_m$  and is defined as:

$$Z_m = A_m(X) = \lambda_m(X) \quad (1.7)$$

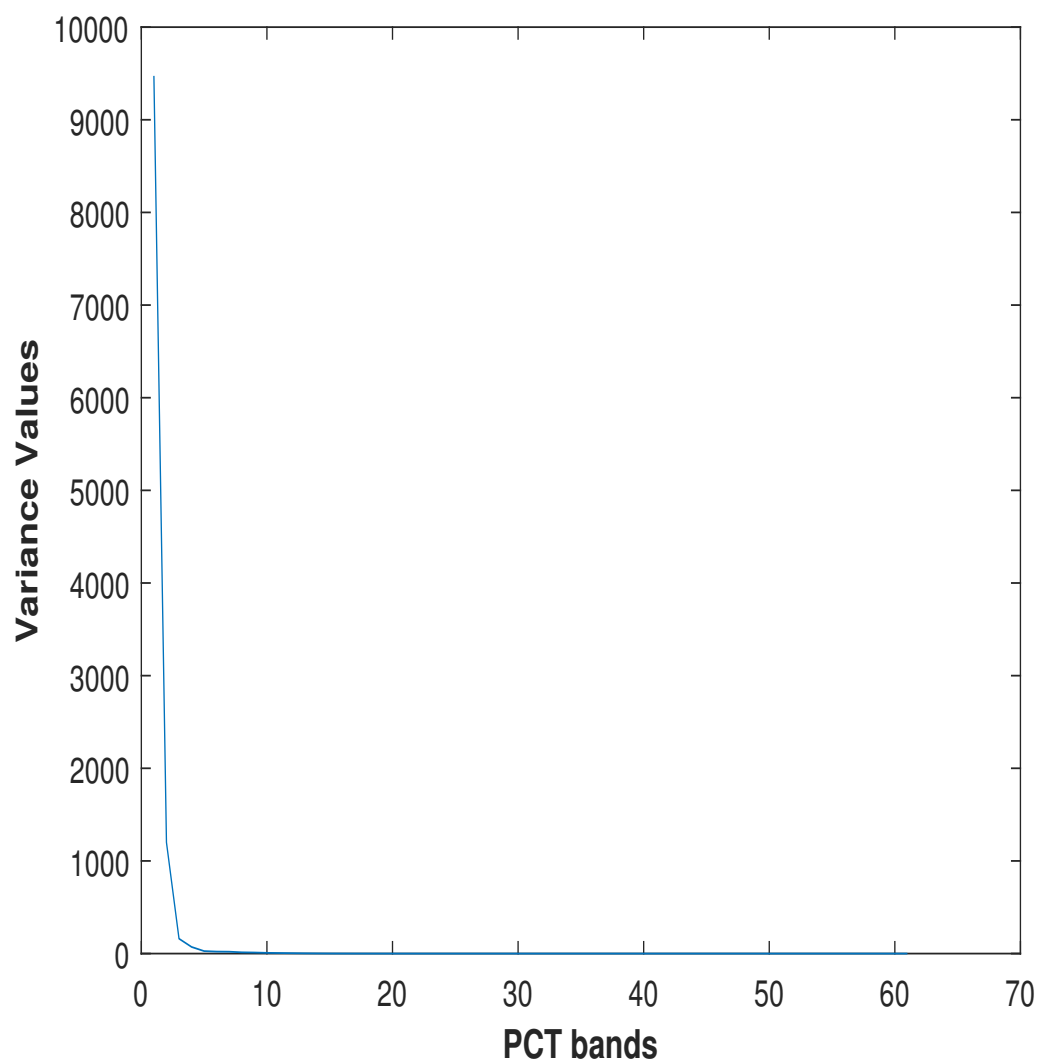
To minimize the computational effort and time while retaining maximum information, the first seven eigenvalues were selected, which corresponded to bands having high information content (Fig 1.4). By substituting  $m = 7$  in equation (1.7), equation (1.8) to obtain the PCT of the vector  $X$  was obtained.

$$Z_7 = A_7X \quad (1.8)$$

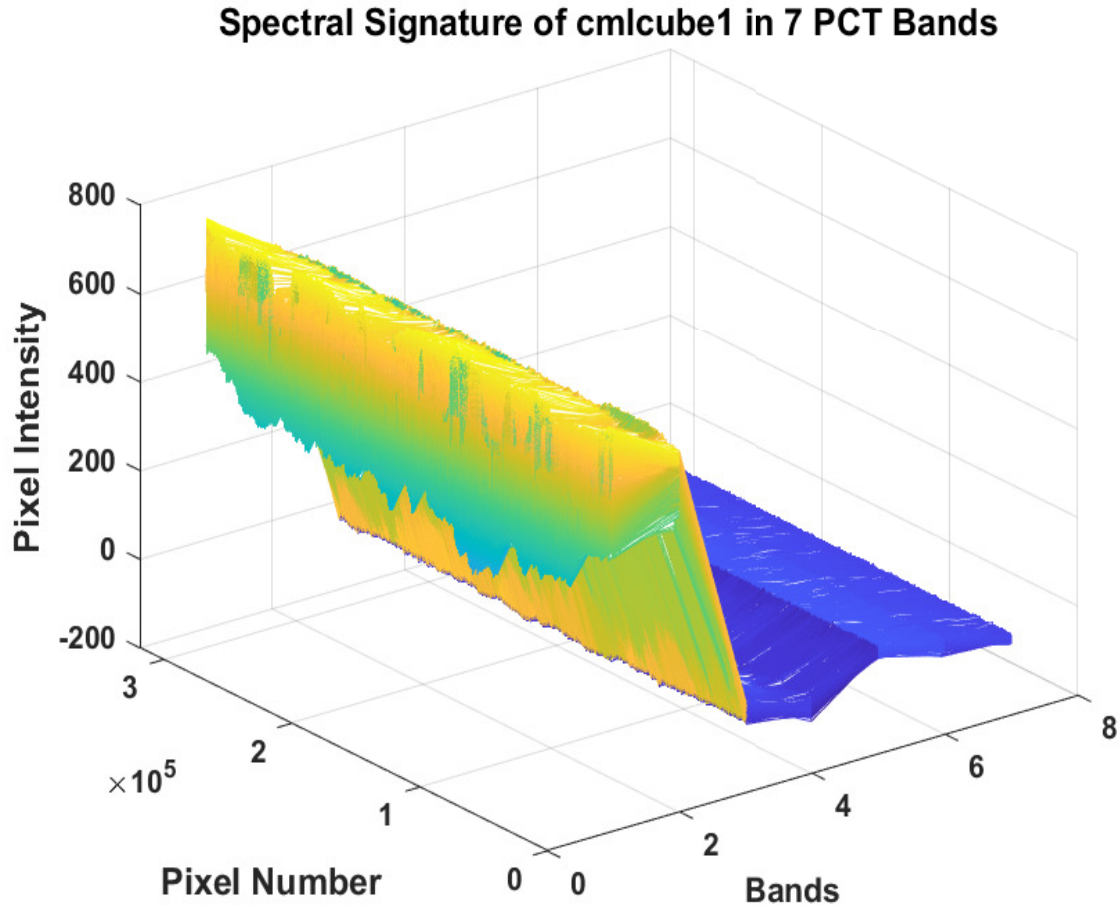
By taking the inverse of  $Z_7$ , the vector  $X_7$  can be obtained from the vector  $X$ .  $X_7$  contains information from the first seven high information containing bands of  $X$ . Therefore, the transformed hyperspectral image cube in the first seven dimensions,  $X_7$  is defined as given by equation (1.9).

$$X_7 = A_7^{-1}(Z_7) \quad (1.9)$$





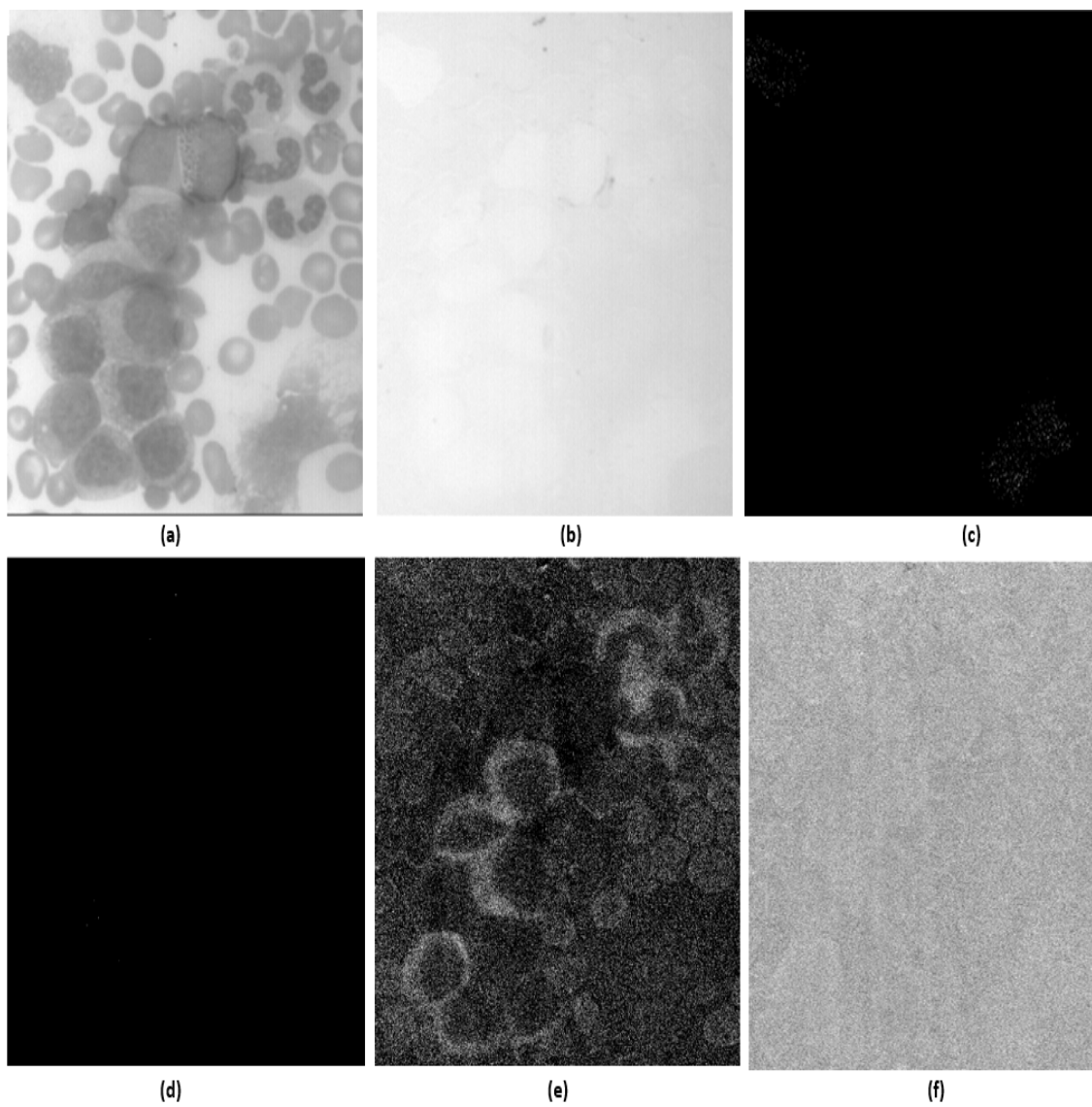
**Figure 1.8** The graphical depiction of reduction in variance values of the hyperspectral image cube (the RGB image is shown in Fig. 1.2.) corresponding to higher PCT bands.



**Figure 1.9** Spectral signature of first hyperspectral cube in the first seven bands after applying PCT.

The pixel values in the first seven bands having maximum information for all the 318516 pixels of the first hyperspectral cube, 'cmlcube1', was plotted as a 3-D mesh (Fig. 1.9).

The images of the first unsplit hyperspectral image cube, 'cmlcube1' in the first 7 PCT wavelength bands is illustrated in Fig. 1.10. It is observed that the data set corresponding to lower PCT wavelength bands have greater information content than the data set corresponding to the higher PCT wavelength bands when arranged in decreasing order of their eigenvalues.



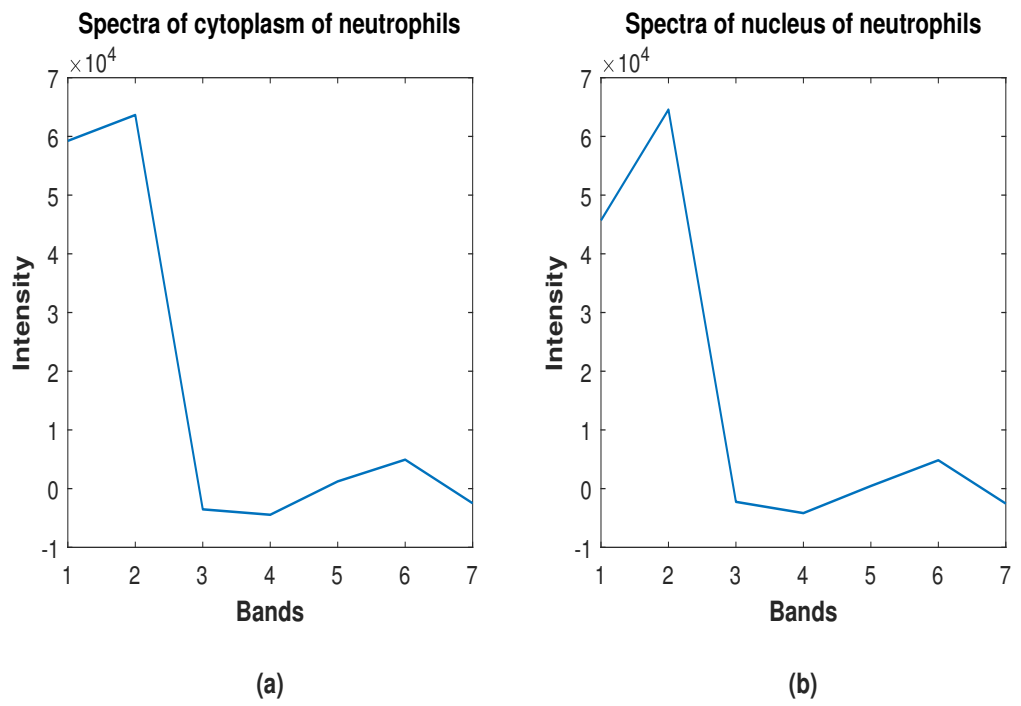
**Figure 1.10** The images in the first six wavelength bands of the hyperspectral image cube in descending order given from (a) to (f).

### 1.4.3 Training Pixels

Since the classification methods explored as part of the thesis were supervised classification methods, it was essential to earmark training pixels. These training pixels were used to train the classification algorithms for classifying the unknown test samples. A robust classification algorithm can achieve better results with lesser number of training pixels as compared to a non-robust algorithm. In order to demonstrate the strength of the classification algorithms being proposed as part of this thesis, minimum quantity of training pixels were used to train the classification algorithms.

We had 13 sample cubes. For implementing classification on 13 cubes, 57 fold cross validation technique was adopted to split each of the 12 cubes to 4 equal folds and one of the cubes to 9 equal folds. Each fold had a different size. For the purpose of training, three sub-image cubes out of the total 57 sub-image cubes were identified after applying PCT. Four sets of pixels depicting the neutrophils were identified from the nucleus and the cytoplasm regions of these three image sub cubes. A subset of  $10 \times 11$  pixels each for the nucleus and the cytoplasm of the four sets of training pixels was selected for the purpose of training. The spectra of the cytoplasm and the nucleus of the CML neutrophils are graphically plotted in Fig. 1.11a and 1.11b.

Wang et. al. [47] had used the 3D ResNeXt algorithm comprising of a convolutional layer and three consecutive residual blocks along with 3-D attention model of a python-based deep convolution neural network package (called deep hyper) to classify white blood cells. However, they used 70% ( $n = 151$ ) of samples for training and 15% ( $n = 32$ ) of samples for testing as compared to the present study where we used 3.5% ( $n = 2$ ) of samples for training and 96.5% ( $n = 55$ ) of samples for testing in all our algorithms. This underlines the inherent strength of classification methods proposed as part of this thesis.



**Figure 1.11** Spectral signature of (a) the cytoplasm and (b) the nucleus part of CML neutrophils in the seven maximum information bands after applying PCT.

#### 1.4.4 Classification Techniques

The classification techniques explored in the research work are supervised algorithms and are based on the features extracted from the training pixels.

The first two methods are based purely on statistical distances in multidimensional space i.e., the Euclidean distance method and the Mahalanobis distance method [21,27,48]. The Euclidean distance method computes the shortest distance between the test and the training samples in multidimensional space. Mahalanobis distance method compensates the variance of the target data distribution while computing the shortest distance between the test and the training samples [37,44]. While the Euclidean distance was superior when it came to the sensitivity (81.8%) measure of CML detection, the Mahalanobis distance was better for specificity (89.1%) and accuracy (82.5%) measures.

To improve upon the specificity and accuracy values, the third and the fourth methods were proposed based on calculation of the spectral angle in multidimensional space i.e., the Spectral Angle Mapping (SAM) method and computation of the spectral-spatial gradient value of the hyperspectral cube along all the three spatial dimensions i.e., the 3-D Spectral Gradient Mapping (3-D SGM) method. SAM is an existing classification method widely used in remote sensing [28]. 3-D SGM is a novel algorithm which improved the specificity and the accuracy values to 97.7% and 84.2% respectively. However, it produced very low sensitivity of 38.5% [45].

To improve the sensitivity values without compromising the specificity and accuracy values, the fifth and sixth methods were proposed. These methods were based on the Probability Distribution Function (PDF) of the samples. The fifth method is a merged method using the Frequency Domain Normal PDF Matching (FDNPM) metric along with Mahalanobis Distance Method to improve the sensitivity without compromising the specificity and accuracy values. FDNPM is based on computing the difference in the frequency domain of the probability distributions of the test and the training samples after modelling them based on normal PDF. It improved the

sensitivity to 63.6%. To further improve the sensitivity without degrading specificity and accuracy, the sixth method i.e., the Kurtosis Compensated Euclidean Distance (KCED) method was proposed to use the kurtosis to measure the spread of the probability distribution of the training sample along a given dimension or band. It then computes a distance metric by compensating the euclidean distance of the test sample from the distribution of the training sample. Out of these two methods, KCED was found to be overall better at classification of CML neutrophils from hyperspectral data sets and has sensitivity, specificity and accuracy values of 81.8%, 93.5% and 91.2% respectively.

#### 1.4.5 Evaluation Parameters

To have a standardised and unbiased evaluation, certain established measurement parameters were adopted like sensitivity, specificity, accuracy and likelihood ratios [47-50]. The effectiveness of the proposed methods have been compared and evaluated by using these measurement parameters. A classification procedure for a test can be characterized by three parameters, i.e., sensitivity, specificity and accuracy. In the present study, the sensitivity of the classification procedure examines its ability to detect CML neutrophils and the specificity of the classification procedure tests its ability to detect healthy neutrophils. The accuracy of the classification procedure examines its ability to differentiate CML from healthy neutrophils [47-51]. The sensitivity, specificity and the accuracy of an algorithm is determined by the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) classification values [47-48]. The positive likelihood ratio [LR(+)] depicts the ratio of the probability of a person having CML and testing positive, to the probability of a person not having CML and testing positive. The negative likelihood ratio [LR(-)] is the ratio of persons having the disease and testing negative to the persons not having the disease and testing negative [49-52].

## 1.5 Motivation

Leukemia is a type of blood cancer that starts in the bone marrow. It causes faster cell division resulting in an abnormal increase in white blood cells (WBCs) and a decrease in the number of red blood cells. Chronic myeloid leukemia (CML) is a specific type of leukemia in which the abnormal blood cells are mature. It is one of the most common blood cancers and takes a longer time to develop. It affects adults more than children and the risk of developing CML increases after the age of 65. The presence of leukemia cells is characterized by an unusual increase in the number of white blood cells [21]. Neutrophils are the most abundant (60 – 70%) cell type in blood circulation. It performs an important role in the innate immunity by providing the first line of defense. The symptoms associated with CML include anaemia (weakness, breathlessness and fatigue), leukopenia (shortage of normal white blood cells resulting in frequent infections), neutropenia (shortage of neutrophils resulting in reduced ability to fight bacterial infections), thrombocytopenia (shortage of blood platelets resulting in eas bleeding and bruising that manifests in frequent nose bleeds and gum bleeds) and thrombocytosis (excessive abnormal platelets which suppresses the normal platelets and result in frequent bleeding). CML is difficult to diagnose in the initial stages. Once it shows symptoms and gets detected, it is difficult to cure as compared to acute leukemia.

Diagnostic tests for CML include complete blood cell count (CBC) obtained from blood samples, bone marrow aspiration followed by biopsy, genetic tests to detect philadelphia chromosome or BCR-ABL gene, cytogenetics or cell karyotyping, Fluorescent in situ hybridization (FISH) to identify BCR-ABL gene in chromosomes, polymerase chain reaction test on the blood samples or bone marrow samples and imaging tests like computer tomography (CT) scans or ultrasound imaging of lymph nodes. However, all these diagnostic tests are either time consuming or display low values of sensitivity, specificity and accuracy. Therefore, there was need to devise an automated faster technique with a high degree of sensitivity, specificity and ac-



curacy. The routine diagnosis for leukemia is based on the visual assessment of a stained blood smear sample using a light microscope. Advances in medical imaging technologies have revolutionized health care delivery globally by making diagnostics more efficient through automation, and goes beyond the limitations of the human eye. One such technology is the hyperspectral imaging system. Hyperspectral imaging was first used in the field of remote sensing in which the hyperspectral sensor was placed on an aerial platform [24-25,42]. The high spectral resolution of the images compensated for the low spatial resolution of the large target areas during analysis.

There is tremendous potential for non-invasive diagnosis of diseases with the use of hyperspectral imaging into the field of medical applications [41]. While the human eye is limited to sensing electromagnetic radiations with wavelengths in the visible range (400–700 nm), hyperspectral sensors provide images of biological samples over multiple near contiguous wavelength bands [21,20,53]. The high spectral and spatial resolutions of hyperspectral images generate a large multi-layered stack where each spectral frequency represents an image layer. The data needs to be processed to extract useful information [17,42]. Current hyperspectral systems capture images with a band gap as low as 10 nm. The specific spectral signatures can be exploited to identify target pixels in sample classification problems [24-25].

## 1.6 Objectives

The objectives of this dissertation are as follows:

- To demonstrate the use of hyperspectral images for classifying CML.
- Explore the use Principal Component Transform (PCT) to achieve reduction in dimensions of the hyperspectral image cube without compromising information content and improve computational efficiency.
- To develop a CML detection method which provides high sensitivity, specificity and accuracy values along with high positive likelihood ratio and low negative

likelihood ratio.

- To achieve better classification results by using minimum training pixels.
- To explore use of multi-classifier algorithms to achieve better results.
- To develop high efficiency classification algorithms using less computation resources and time.

In order to achieve the objectives as given above, the dissertation aims to develop the following classification methods and techniques:

- Development of a CML detection method based on multi-variate Euclidean and Mahalanobis distances applied in hyperspectral space and evaluate it.
- Development of a diagnostic method for CML using SAM and 3-D SGM in hyperspectral space and evaluate it
- Development of a multi-classifier by suitably merging frequency domain normal PDF matching metric (FDNPM) with Mahalanobis distance and use it for classifying CML cases from blood samples and evaluate it.
- Development of Kurtosis Compensated Euclidean Distance (KCED) and propose its use for classifying CML affected blood samples and evaluate it.

## 1.7 Contributions

The contributions to this thesis are summarized as follows:

- The use of statistical distances like Euclidean distance and Mahalanobis distance computed in hyperspectral space has been developed and compared to draw analysis of the advantages and disadvantages. While the Euclidean distance was superior when it came to the sensitivity (81.8%) measure, the Mahalanobis distance was better for specificity (89.1%) and accuracy (82.5%) measures for detection of CML neutrophils.

- To further improve the measurement parameters, two methods were proposed based on the spectral characteristics inherent to the hyperspectral cube. The first method is the windowed SAM method and the second method is the 3-D SGM method.

In the present study, SAM was implemented over a group of test pixels in reduced bands. This variation of implementation of SAM has been named the Windowed SAM method.

The 3-D SGM method has been proposed to overcome limitations in windowed SAM method by exploiting the inter-pixel and inter-band information content of the hyperspectral cube after carrying out necessary reduction in dimensions. This method not only exploits the spectral information but also uses the inter pixel information to classify ML neutrophils.

The classification carried out using this method was compared with windowed SAM method and analysed. The specificity measure for the algorithms showed that 3-D SGM (97.7%) was superior to Windowed SAM (72.7%). It was also found to be superior to the Mahalanobis distance method (89.1%) in ruling out the presence of disease. The accuracy shown by the 3-D SGM (84.2%) was fairly higher than that of Mahalanobis distance method (82.5%) which showed better ability to distinguish CML neutrophils from healthy neutrophils. Also the high value of positive likelihood ratio indicated that the 3-D SGM was superior in diagnosing the presence of the disease (i.e., positive test for CML) versus Windowed SAM method. However 3-D SGM had a very low sensitivity (38.5%).

- Two different methods based on the Probability Distribution Function (PDF) were proposed to improve the sensitivity, specificity and accuracy values in classifying CML affected neutrophils from hyperspectral images.

The first method is a combination of two methods based on PDF. It proposes a composite metric consisting of two parameters, i.e. the Mahalanobis

distance and the Frequency Domain Normal PDF Matching (FDNPM) parameter. This merged method exploits the benefits of both the Mahalanobis distance method which is computed in the spatial domain and the FDNPM method which capitalises on the information content in the frequency domain

The second method is the Kurtosis Compensated Euclidean Distance (KCED) method. The merged method using the composite metric reduces the likelihood of missing diseased cases with a lower LR(-) value and improves the sensitivity without compromising the high specificity and accuracy values in the Mahalanobis distance method. KCED significantly increases the specificity, accuracy and LR (+) values and reduces the LR(-) value, while maintaining specificity values as obtained in the euclidean distance method. Therefore, overall KCED was found to be superior to the merged method.

## 1.8 Organization of the Thesis

The remaining portion of this thesis is organized in the following way:

- In chapter 2, the use of SAM has been demonstrated on the hyperspectral image cubes. The lacunae in the method was identified. Accordingly, the 3-D SGM method was proposed to exploit spectral as well as spatial information contained in the hyperspectral cubes for classification.
- In chapter 3, the use of statistical distances in hyperspectral space has been explored for classification. In this chapter multi-dimensional Euclidean distance and Mahalanobis distance has been used to measure the similarity between training and test pixels of hyperspectral images to classify CML affected neutrophils.
- In chapter 4, a multiclassifier approach towards classification of CML neutrophils from hyperspectral images has been attempted. In this chapter, frequency domain normal PDF matching has been used along with Mahalanobis

distance to achieve classification of CML affected neutrophils. The results obtained have been compared and discussed with the other results. To achieve frequency domain analysis, fourier transform of the PDF was undertaken before attempting to match it to the normal PDF of the training samples.

- In chapter 5, the use of kurtosis compensated distance parameter derived from the multi-dimensional probability distribution has been proposed for evaluating the similarity of the training and the test pixels of hyperspectral cube to classify CML affected neutrophils. The results obtained were measured using metrics like sensitivity, specificity, accuracy and likelihood ratios. The results were compared with results obtained using other methods and discussed in detail.
- Finally, the whole work is concluded in chapter 7, The scope for further research work in future has also been discussed in this chapter.



## Chapter 2

# Spectral Angle Mapping (SAM) and 3-D Spectral Gradient Mapping Method (3-D SGM)

### 2.1 Introduction

Spectral angle mapping is a tested algorithm which has been used extensively in classification of linearly mixed datasets [53]. It has been used for classifying low spatial resolution remotely sensed images [16]. However, for high spatial resolution images, SAM needs to be implemented in a windowed manner to obtain optimised results. In this chapter, the use of spectral angle mapping has been carried out by computing the spectral angle between the test and the training pixels by considering the pixel intensities in each wavelength band as a spectral dimension to classify CML neutrophils from healthy neutrophils for diagnosis of CML blood samples. The implementation of SAM has been done by taking windows of  $2 \times 2$  pixels, i.e. Windowed SAM method. The shortcomings in the use of windowed SAM method for classifying hyperspectral images has been discussed. To overcome the shortcomings the use of 3D-SGM has been proposed and compared with Windowed SAM method. This new algorithm enables extraction of additional inter-band information of the

inter-pixel differences and the inter-pixel information of the inter-band differences, which helps in better classification of high spatial and high spectral resolution hyperspectral images.

## 2.2 Spectral Angle Mapping (SAM)

SAM is a supervised classification method in which the decision is made based on learning from training data [21,24]. Based on linear mixture modeling [22,33], every pixel of the hyperspectral cube can be considered to be a multidimensional vector. In this kind of modeling, the intensity value of each pixel can be considered to be a linear sum of the scaled intensities of the individual end members. The end members are vectors of unit magnitude along the different dimensions being considered [20,22]. Here the end members are unit vectors in the various wavelength bands. Since reflectance along each wavelength is independent of the reflectance along the other wavelengths, they can be considered orthogonal. The spectral angle  $\alpha$  between any pixel at the coordinates  $(i, j)$  in the data set and the training pixels is computed based on the formula given at equation (2.1) [21].

$$\alpha(i, j) = \cos^{-1} \left[ \frac{\sum_{k=1}^{nb} x_k y_k}{\sqrt{\sum_{k=1}^{nb} x_k^2} \sqrt{\sum_{k=1}^{nb} y_k^2}} \right] \quad (2.1)$$

where,

$nb$ - Total number of bands (Here we have considered a reduced number of seven bands).

$x_k$ - Spectrum of the pixel under consideration over the reduced number of bands.

$y_k$ - Spectrum of the training pixels over the reduced number of bands.



### 2.2.1 Implementation of SAM: Windowed SAM

SAM was implemented in this study as Windowed SAM using a window cube of size of  $2 \times 2 \times 7$  intensity values corresponding to four pixels over seven wavelength bands. In the Windowed SAM method, the four  $2 \times 2$  training pixel groups were averaged to form a square having averaged spectrum of the  $2 \times 2$  training pixels over the seven reduced wavelength bands. This  $2 \times 2 \times 7$  window cube of training pixels was moved from one corner of the padded hyperspectral cube so as to cover the entire image. At each position of this window, the pixel-wise spectral angles were computed between all the four test pixels from their corresponding four training pixels over the reduced wavelength bands. The sum of these angles was calculated to obtain the cumulative spectral angle as given at equation (2.2). A particular pixel was detected as a CML neutrophil if the cumulative spectral angle in respect to the pixel,  $\beta$  was found to be less than the threshold angle  $\theta$ , as given in equation (2.3):

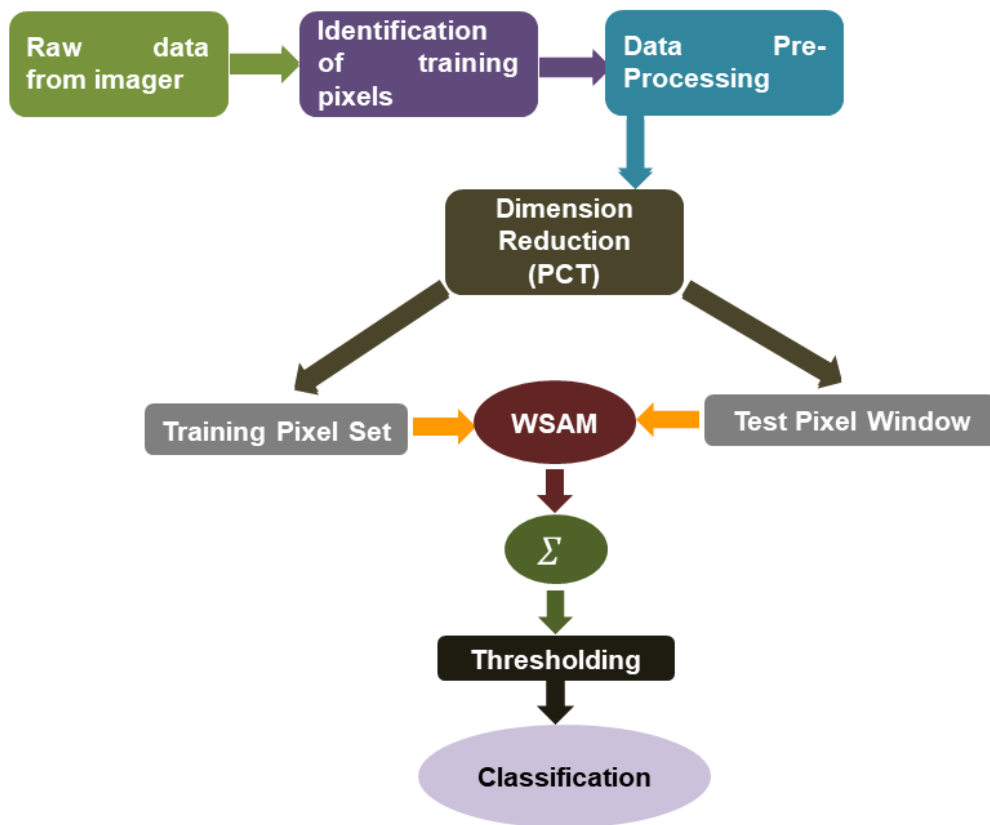
$$\beta(i, j) = \alpha(i, j) + \alpha(i, j + 1) + \alpha(i + 1, j) + \alpha(i + 1, j + 1) \quad (2.2)$$

$$Detection \implies \beta < \theta \quad (2.3)$$

A schematic flow chart showing the sequence of steps required to be executed for implementing Windowed SAM algorithm on hyperspectral dataset for the purpose of classification is given at Fig. 2.1.

### 2.2.2 Shortcomings in SAM

SAM is a powerful detection algorithm that exploits the information content of the spectral bands of a particular pixel, which can be used for classification. It has been widely used in the low resolution remotely sensed hyperspectral images of the earth surface [21,45]. However for high resolution biomedical hyperspectral images



**Figure 2.1** Flowchart showing the implementation of Windowed SAM method for classification of hyperspectral image.

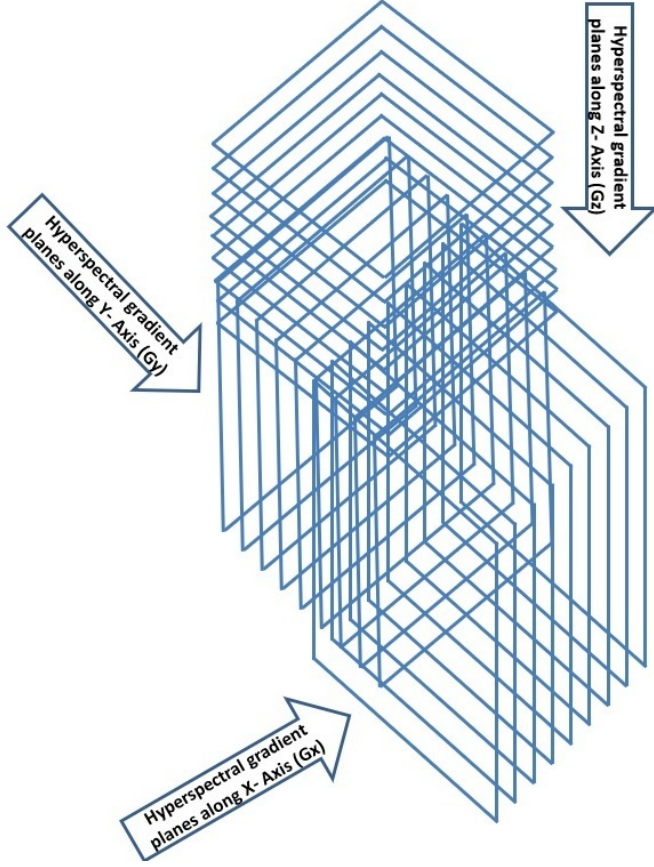
like blood samples where the similarity between the different types of blood cells is high, the spectral information is not sufficient for correct detection. Therefore in order to further extract information out of the hyperspectral cube, it is necessary to extract inter-band information of the inter-pixel differences and to extract inter-pixel information of the inter-band differences. This has been achieved by 3-D Spectral Gradient Mapping (3-D SGM).

## 2.3 3-Dimensional Spectral Gradient Mapping (3-D SGM)

To yield optimum information out of the hyperspectral cube, we exploited the 3-D gradient information to form a new vector cube of 3-D gradient vectors. In this method, a 3-D hyperspectral gradient vector was built up from the hyperspectral cube. The construct of the hyperspectral gradient vector cube consists of stacked planes in the three orthogonal directions. A 3-D gradient vector was computed for each point of intersection of the stacked planes. Thus, every point in the hyperspectral gradient cube was a part of three orthogonal gradient planes and hence has three mutually independent orthogonal gradient values. Therefore, each point of the hyperspectral gradient cube was denoted by a 3-D vector.

### 2.3.1 Gradient Planes

The gradient of the planes along each direction was computed. So for a  $x \times y \times 7$  hyperspectral cube along reduced dimensions, there were 7 gradient planes of size  $x \times y$  denoted as  $G_z$ ,  $x$  gradient planes of size  $y \times 7$  denoted as  $G_x$  and  $y$  gradient planes of size  $7 \times x$  denoted as  $G_y$  as listed in equations (2.4), (2.5) and (2.6)



**Figure 2.2** Gradient planes perpendicular to the X, Y and Z axis giving each pixel as a gradient vector with three components corresponding to the three planes.

respectively. Here the maximum value of Z was 7.

$$G_x = \nabla(Y - Z \text{ hyperspectral plane}) \quad (2.4)$$

$$G_y = \nabla(Z - X \text{ hyperspectral plane}) \quad (2.5)$$

$$G_z = \nabla(X - Y \text{ hyperspectral plane}) \quad (2.6)$$

A pictorial representation of the three gradient planes formed out of the hyperspectral cube is shown in Fig. 2.2.

### 2.3.2 3-D Spectral Gradient Cube and 4-D Pixel Value of the Spectral Gradient Cube

The new spectral gradient cube formed is of the same size as the hyperspectral image cube of reduced dimensions. Each pixel location of this cube corresponds to a 3-D vector, with each dimension corresponding to the gradient value along the three planes, viz,  $G_x$ ,  $G_y$  and  $G_z$ . So, effectively each pixel was represented by a 4-D vector, where the first three dimensions specify the location of the pixel in the spectral gradient cube and the fourth dimension specifies the value of the pixel along the 3 gradient planes, and can be called a vector dimension. Therefore, the vector denoting the pixel located at  $(x, y, z)$  in the hyperspectral cube, with the fourth dimension specifying its value along the gradient planes  $G_x$ ,  $G_y$ ,  $G_z$ , was denoted by  $g$ . The values of the vector  $g$  along the three gradient planes, i.e. the  $Y - Z$ ,  $Z - X$  and  $X - Y$  planes are given by equations (2.7), (2.8) and (2.9) respectively.

$$g(x, y, z, 1) = G_x(x, y, z) \quad (2.7)$$

$$g(x, y, z, 2) = G_y(x, y, z) \quad (2.8)$$

$$g(x, y, z, 3) = G_z(x, y, z) \quad (2.9)$$

### 2.3.3 Spectral Gradient Cube of Training Pixels

For classification of the hyperspectral image using 3-D SGM, it was necessary to compute the 3-D spectral gradient cube out of the training pixels with each cell of the cube depicting a 3-D gradient vector. A 57 fold cross validation technique as described at paragraph 1.4.3 was implemented to test the classification algorithm. Four thousand three hundred training pixels were identified from the first six CML sample cube folds belonging to four different neutrophils. These groups of training

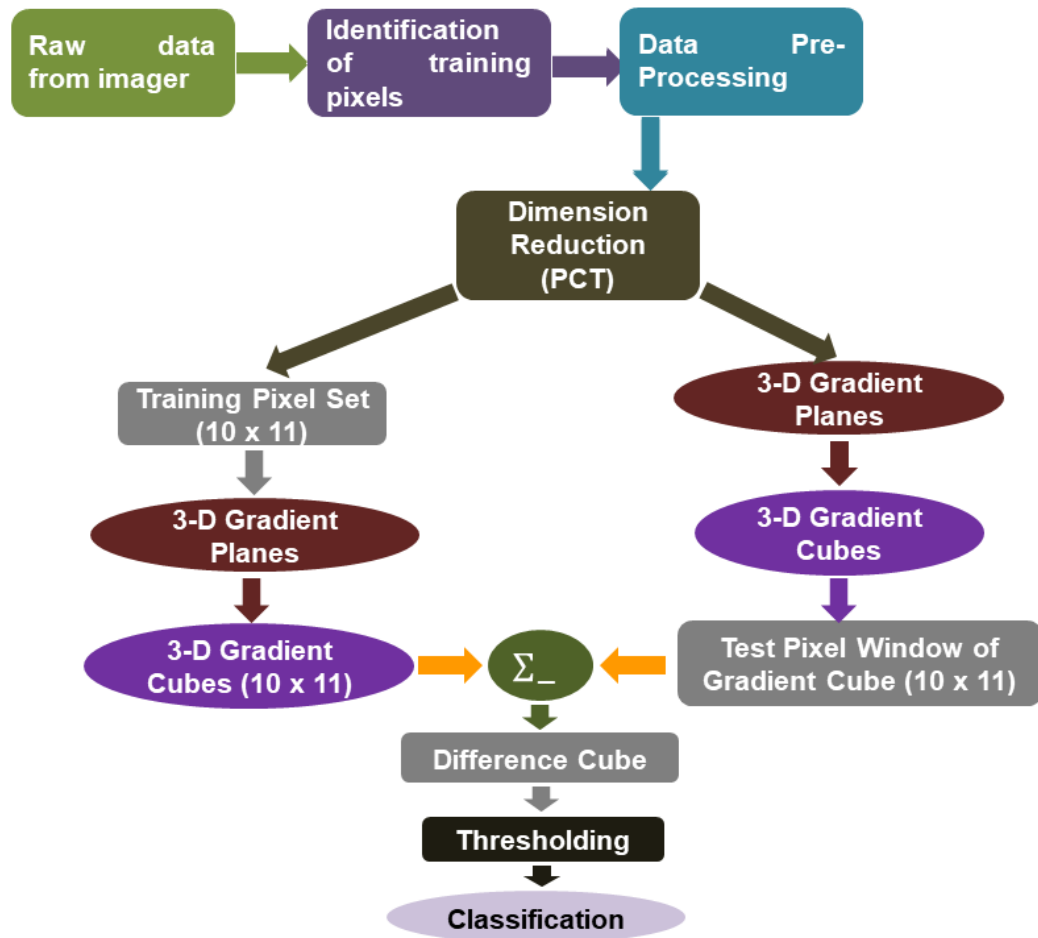
pixels were averaged out to groups of four vector cubes of size  $10 \times 10 \times 7$ , in the first seven bands having high information content. After having formed the four spectral gradient cubes with each coordinate depicting a cell containing a 3-D gradient vector, they were averaged vector-wise, to form a  $10 \times 10 \times 7$  composite spectral gradient cube of training cells with a 3-D vector in each cell.

### 2.3.4 Implementation of 3-D SGM

In this method, the composite spectral gradient cube formed out of the training pixels was moved from one corner to another of the padded composite spectral gradient cube. At each location, the difference in spectral gradient was calculated and a spectral gradient difference cube was computed with each cell of the cube containing the difference of the 3-D spectral gradient vectors of the test and the training hyperspectral cubes (Fig 2.1). At each position of this window, the coordinate-wise 3-D spectral gradient vector value was averaged over the three dimensions to obtain a composite spectral difference cube. The cube for the pixels representing the cytoplasm and nucleus was computed and plotted as an image, each respectively. These two images clearly gave the shape of the CML neutrophils with high radiance values. Further, based on a defined threshold of 0.9 times the maximum 3-D spectral gradient difference vector in each dimension, the detected pixels were shaded white and plotted. By comparing the above three images, any pixel having high radiance values in all the three images was detected as positive for CML neutrophil.

A schematic flow chart showing the sequence of steps required to be executed for implementing Windowed SAM algorithm on hyperspectral dataset for the purpose of classification is given at Fig. 2.3.

The RGB image of the fourth hyperspectral sub-cube of the third hyperspectral cube obtained after taking only the wavelengths corresponding to red, green and blue colors, is illustrated in Fig 2.4a. After applying 3-D SGM on the ibid cube, the complement of the composite spectral gradient difference cube, with the pixels



**Figure 2.3** Flowchart showing the implementation of 3-D SGM method for classification of hyperspectral image.

that are supposed to represent the cytoplasm shown with high radiance values is depicted in Fig 2.4b. The composite spectral gradient difference cube, with the pixels that are supposed to represent the nucleus shown with high radiance values is depicted in Fig 2.4c. The binary image formed after combining both the images and thresholding is depicted at Fig 2.4d. The detected pixels are shown in white. The results showed that the algorithm was able to detect CML neutrophils.

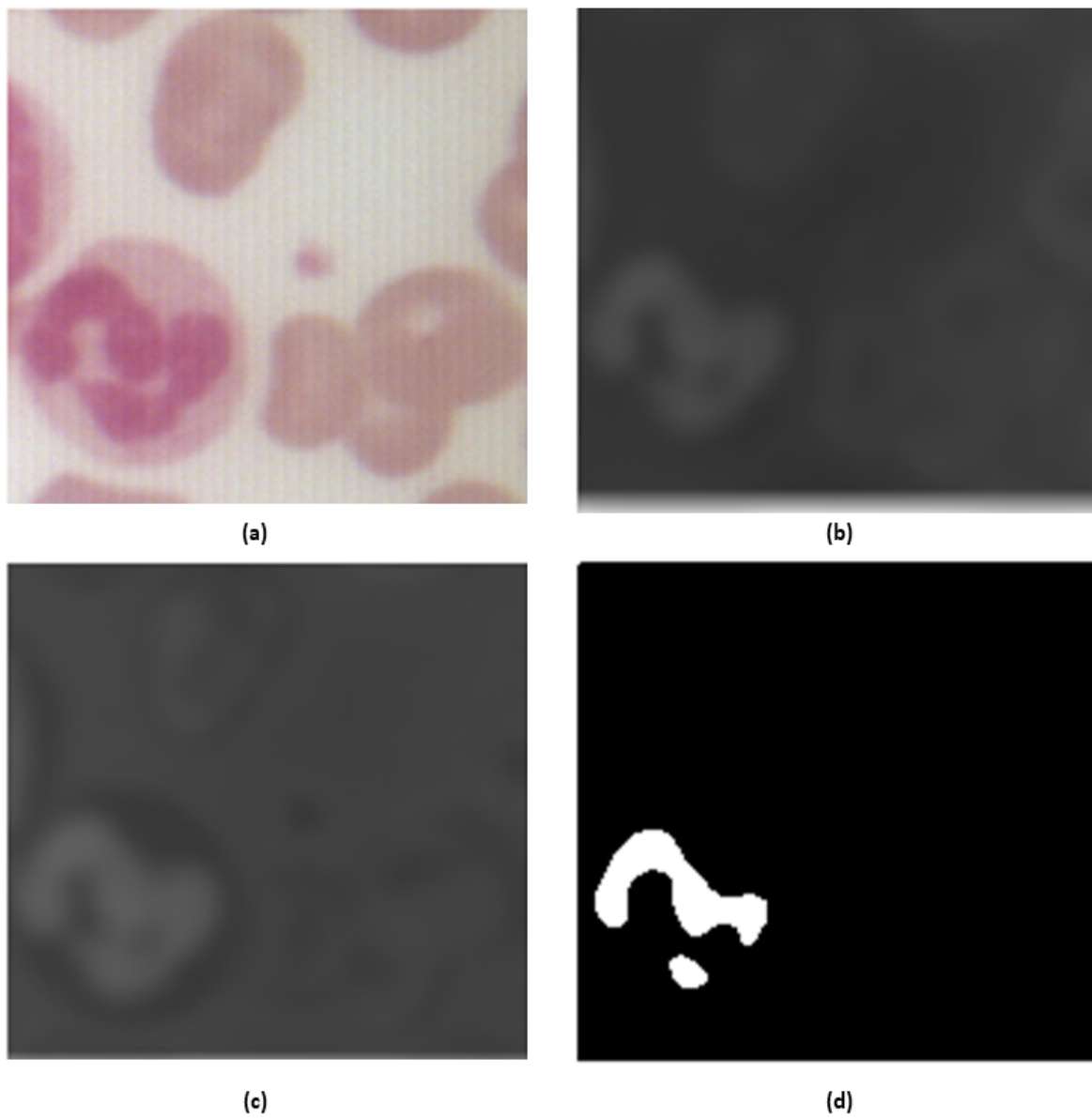
## 2.4 Results

Sensitivity, specificity, accuracy and likelihood ratios (LRs) were used maintaining uniform parameters to measure performance of SAM and 3-D SGM. While sensitivity is the ability of a test to identify a disease condition correctly, specificity is the test's ability to exclude a disease condition correctly. The accuracy of the classification procedures examines its ability to differentiate CML from healthy neutrophils. The LR indicates how many times more (or less) likely a test result (either positive or negative) is to be found in a diseased compared with non-diseased person.

### 2.4.1 Sensitivity, Specificity and Accuracy

The sensitivity, specificity and accuracy measures were computed from, the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) classification values as tabulated (Table 2.1). Sensitivity brings out the percentage of correct positive diagnosis cases out of the total positive samples. With a higher sensitivity value, Windowed SAM was better at ruling in a diseased case when compared to 3-D SGM. Specificity brings out the percentage of correctly ruling out the presence of disease out of the total non-diseased samples [47,51]. With a higher specificity value, 3-D SGM was found to better at ruling out the presence of the disease compared to Windowed SAM. As the sensitivity and specificity measures





**Figure 2.4** Detection of CML neutrophil in the sample *cmlcube3s\_2.2*. (a) The RGB depiction of the CML cube; (b) the composite spectral gradient difference cube of cytoplasm and (c) nucleus; (d) the binary image showing the detected pixels for neutrophils.

**Table 2.1** Sensitivity, specificity and accuracy of Windowed SAM versus 3-D SGM algorithms.

Algorithm	Classification				Sensitivity <sup>e</sup> (%)	Specificity <sup>f</sup> (%)	Accuracy <sup>g</sup> (%)
	TP <sup>a</sup>	TN <sup>b</sup>	FP <sup>c</sup>	FN <sup>d</sup>			
<b>Windowed SAM</b>	8	32	12	5	61.5	72.7	70.2
<b>3-D SGM</b>	5	43	1	8	38.5	97.7	84.2

<sup>a</sup> TP represents True Positive

<sup>b</sup> TN represents True Negative

<sup>c</sup> FP represents False Positive

<sup>d</sup> FN represents False Negative

<sup>e</sup> Sensitivity is given by  $TP/(TP+FN)$

<sup>f</sup> Specificity is given by  $TN/(TN+FP)$

<sup>g</sup> Accuracy is given by  $(TP+TN)/(TP+TN+FP+FN)$

are inversely proportional, this outcome between and within these algorithms were as expected. The sensitivity measure between both algorithms showed that Windowed SAM was better in ruling in a disease state because there was a 38.5% chance of missing the diagnosis versus 3-D SGM which has a 61.5% chance of missing the diagnosis. The specificity measure for the algorithms showed that 3D-SGM was superior to Windowed SAM in ruling out the disease condition because with the former algorithm, there was only be 2.3% chance of missing the diagnosis versus 27.3% with the later.

## 2.4.2 Likelihood Ratios

Likelihood ratios were also used as parameters for evaluation the classification methods [48-49]. The positive likelihood ratio [LR(+)] indicated that the 3-D SGM method has a greater diagnostic weight in arguing towards the presence of the disease (i.e., positive test for CML) as compared to Euclidean distance, Mahalanobis distance and Windowed SAM methods. On the contrary, the lower negative likelihood ratio [LR(-)] of Windowed SAM method made it better at ruling out the presence of disease (i.e., negative test for CML) versus 3-D SGM (Table 2.2). How-

**Table 2.2** Likelihood ratios of Windowed SAM and 3-D SGM algorithms.

Algorithm	LR(+) <sup>a</sup>	LR(-) <sup>b</sup>
Windowed SAM	2.26	0.53
3-D SGM	16.70	0.63

<sup>a</sup> LR(+) is the Positive Likelihood Ratio and is calculated as Sensitivity/(1-Specificity)

<sup>b</sup> LR(-) is the Negative Likelihood Ratio and is calculated as (1-Sensitivity)/Specificity

ever, the Euclidean distance and the Mahalanobis distance methods were better at ruling out the presence of disease with much lower negative likelihood ratios [LR(-)] at 0.26 and 0.51 respectively.

## 2.5 Discussion

SAM is a powerful detection algorithm that exploits the information content of the spectral bands of a particular pixel, which can be used for classification. This technique had been first implemented for classification of low resolution remotely sensed images in which the similarities between neighbouring pixels is high [59]. However, for high resolution biomedical hyperspectral images, where the similarities between neighboring pixels are high, the spectral information is not sufficient for correct detection. It is necessary to exploit the inter-band and inter-pixel information by extracting the 3-D spectral gradient vector of data set in the hyperspectral cube. Therefore, this novel method of classification has been called 3-D spectral gradient mapping.

The specificity and sensitivity measures are inversely proportional. In this study, the specificity and accuracy values of 3-D SGM method were higher than those of the Windowed SAM method. It was also found to be higher than the Euclidean and Mahalanobis distance methods, which have been discussed in chapter 3. However, the sensitivity value was found to be lower for 3-D SGM method versus all the

other three methods, i.e. Windowed SAM (discussed in present chapter and the Mahalanobis and the Euclidean distance methods discussed in Chapter 3). The detailed sensitivity, specificity and accuracy values have been tabulated in Table 2.1 (for Windowed SAM and the 3-D SGM methods) and Table 3.1 (for the Euclidean and Mahalanobis distance methods). The 3-D SGM method exploits the spectral as well as the gradient information of the cube, which brings out the information content in the differences between hyperspectral gradient cube points or elements. The potency of 3-D SGM method compared to the Windowed SAM lies in the fact that the window used for test and training pixels used in 3-D SGM is  $10 \times 10 \times 7$  whereas the window used for Windowed SAM is  $2 \times 2 \times 7$ . This indicates that 3-D SGM is able to give better classification results at lower granularity than Windowed SAM method. Lower granularity translates to lesser number of shifts of the window across the hyperspectral cube and hence faster algorithm. However due to more number of computations in each cycle, the benefits of lower granularity is offset in 3-D SGM. The 3-D SGM method exploits the spectral as well as the gradient information of the cube, which brings out the information content in the differences between hyperspectral gradient cube points or elements.

The 97.7% specificity measure for 3-D SGM followed by its superior LR(+) value showed its possible potential as a screening test especially for family members of patients, and also for use with remission cases as a positive diagnostic result rules in CML. In cases of remission, as the patient's neutrophil population becomes 'genetically' healthy, the cells would bind the Leishman dye as would a neutrophil from a healthy person. In other words, the spectral signature would be that which resembles a disease-free (i.e., healthy) person.



# Chapter 3

## Statistical Distances

### 3.1 Introduction

The classification algorithms discussed in the previous chapter were based on spectral angle and difference of the spectral gradient vectors of the hyperspectral images. It is possible to compute the distance between two points in hyperspectral space based on the concept of distance calculation in multidimensional space. In this chapter, two types of distances between the test pixel and the set of training pixels was computed and used for classification. The first type of distance was the Euclidean distance which assumed that the set of training pixels were uniformly distributed in all the hyperspectral dimensions and hence could be represented by a single training pixel. The second type of distance was the Mahalanobis distance which accounted for non-uniformity of the distance along different dimensions based on the actual nature of distribution of the training pixels.

#### 3.1.1 Classification using Euclidean Distance

Euclidean distance is the shortest distance between two points or vectors in a single or multidimensional space. Modified Euclidean distance has been used with face recognition algorithms on gray scale images by Wang et al. [60]. However, for a

hyperspectral image with high spectral information content, the Euclidean distance needs to be computed over multiple spectral dimensions to generate the classification metric.

The averaged single training pixel obtained from four sets of  $10 \times 11 \times 7$  training pixels was computed as given at equation (3.1). To classify any test pixel,  $test(i, j)$ , the Euclidean distance,  $dtest_{i,j}$  of that pixel from the averaged single training pixel over the seven reduced bands was computed as per formula given at equation (3.2)[45].

$$TP_{10 \times 11 \times 7} = \left[ \frac{TP_{10 \times 11 \times 7} + TP_{20 \times 11 \times 7} + TP_{30 \times 11 \times 7} + TP_{40 \times 11 \times 7}}{4} \right] \quad (3.1)$$

$$dtest_{e(i,j)} = \left[ \sum_{k=1}^{nb} (test(i, j) - TP_{10 \times 11 \times 7})^2 \right]^{1/2} \quad (3.2)$$

where,

$dtest_{e(i,j)}$  - The multi-dimensional Euclidean distance between the test pixel with coordinates (i,j) and the average of the set of training pixels given by  $TP_{10 \times 11 \times 7}$ .

$nb$  - Total number of bands/ dimensions (after applying PCT, the number of bands were reduced to seven).

The Euclidean distance of the test pixel was computed from the averaged single training pixel over the seven high information containing bands. The maximum inter-pixel euclidean distance represents the maximum variance between the pixels representing neutrophils from patients suffering from CML in terms of Euclidean distance. The maximum inter-pixel euclidean distance for each of the four sets of  $10 \times 11 \times 7$  training pixels is given by equation (3.3).

$$d_{e max} = \frac{\sum_{i=1}^4 d_{e max_i}}{4} \quad (3.3)$$

where,

$d_{e max}$  - The average of the maximum inter-pixel distance of the four sets of  $10 \times 11$  training pixels.

$d_{emax_i}$  - The maximum inter-pixel distance of the  $i^{th}$  group or set of  $10 \times 11$  training pixels.

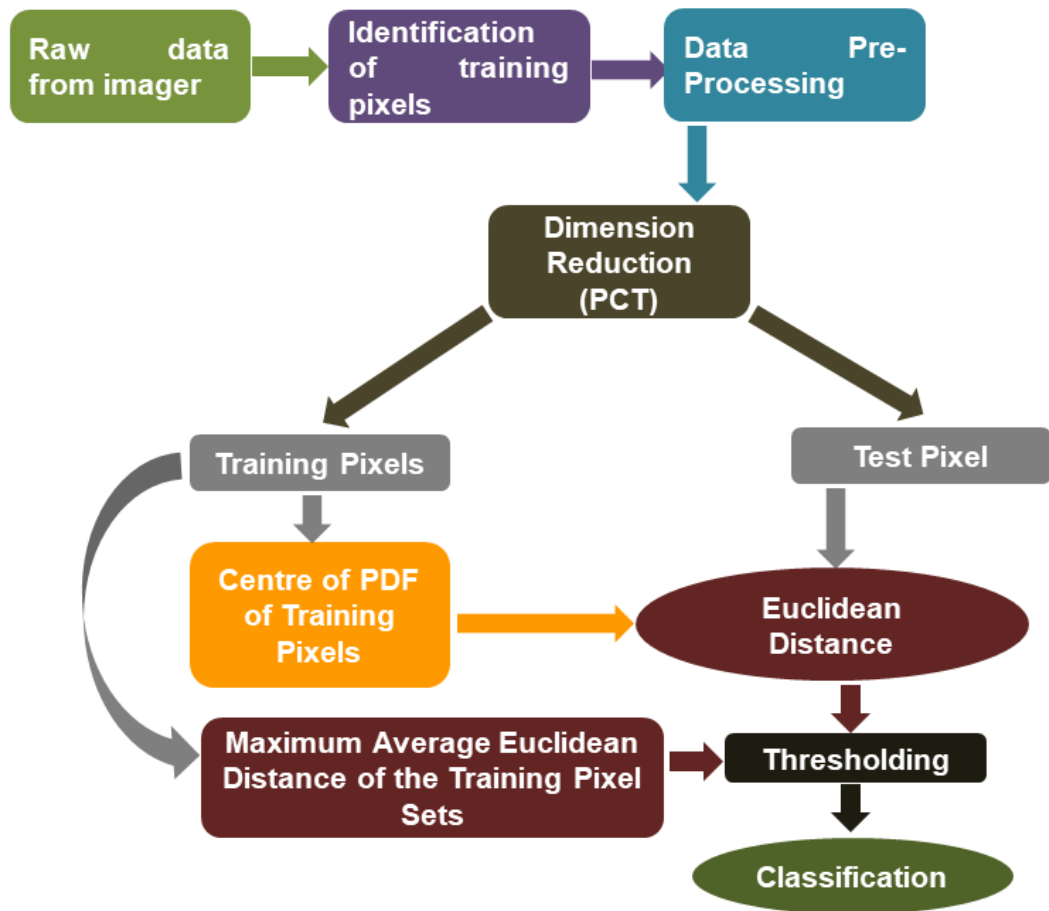
The The test pixel ( $test(i, j)$ ) was classified as a positive CML case if its Euclidean distance, ( $d_{test_{e(i,j)}}$ ) from the single training pixel at the coordinates  $(i, j)$  was found to be less than the largest inter-pixel distance, ( $d_{emax}$ ) of the training pixel set for either the nucleus or cytoplasm of the neutrophil. After highlighting all such cases of positive detection, a decision on classification was taken based on a visual appreciation of the highlighted pixels.

A schematic flow chart showing the sequence of steps required to be executed for implementing Euclidean distance method based classification algorithm on hyperspectral dataset for the purpose of classsification is given at Fig. 3.1.

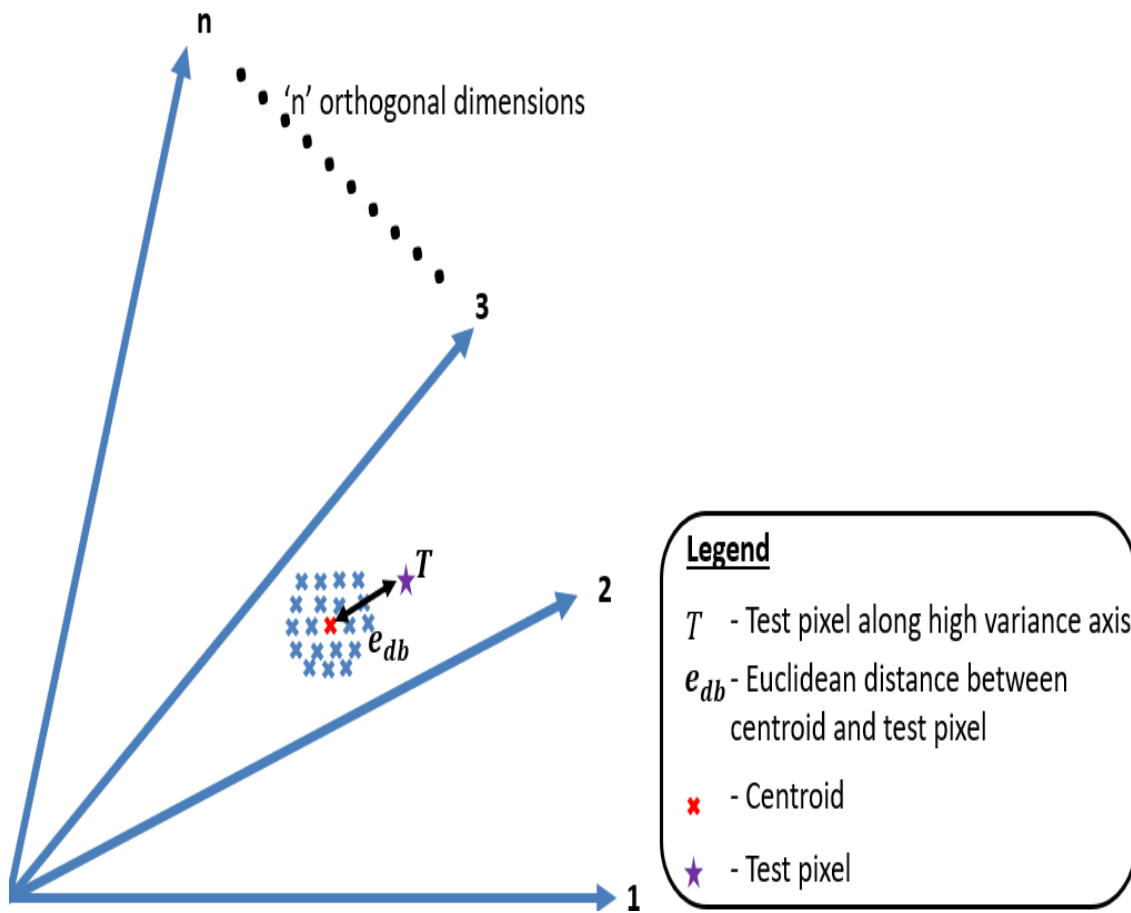
### 3.1.2 Limitations of Euclidean Distance for Classification

Since Euclidean distance is calculated between two points, it is necessary to denote the given distribution of training pixels by a single point. For a balanced or symmetric uncorrelated distribution, considering the centroid as the single point representing the distribution, the threshold can be considered to be the maximum Euclidean distance between the centroid and any pixel of the distribution as shown in Fig. 3.2. However practical distributions may not be a balanced or symmetric distribution. In such cases, the classification of a pixel would depend on the variance of the distribution along each dimension. In Fig. 3.3. even though  $T_a$  and  $T_b$  are at equal Euclidean distance from the centroid of the hyperspectral distribution,  $T_b$  is not a part of the distribution in n-D space. Here Euclidean distance fails to accurately classify the test pixels.

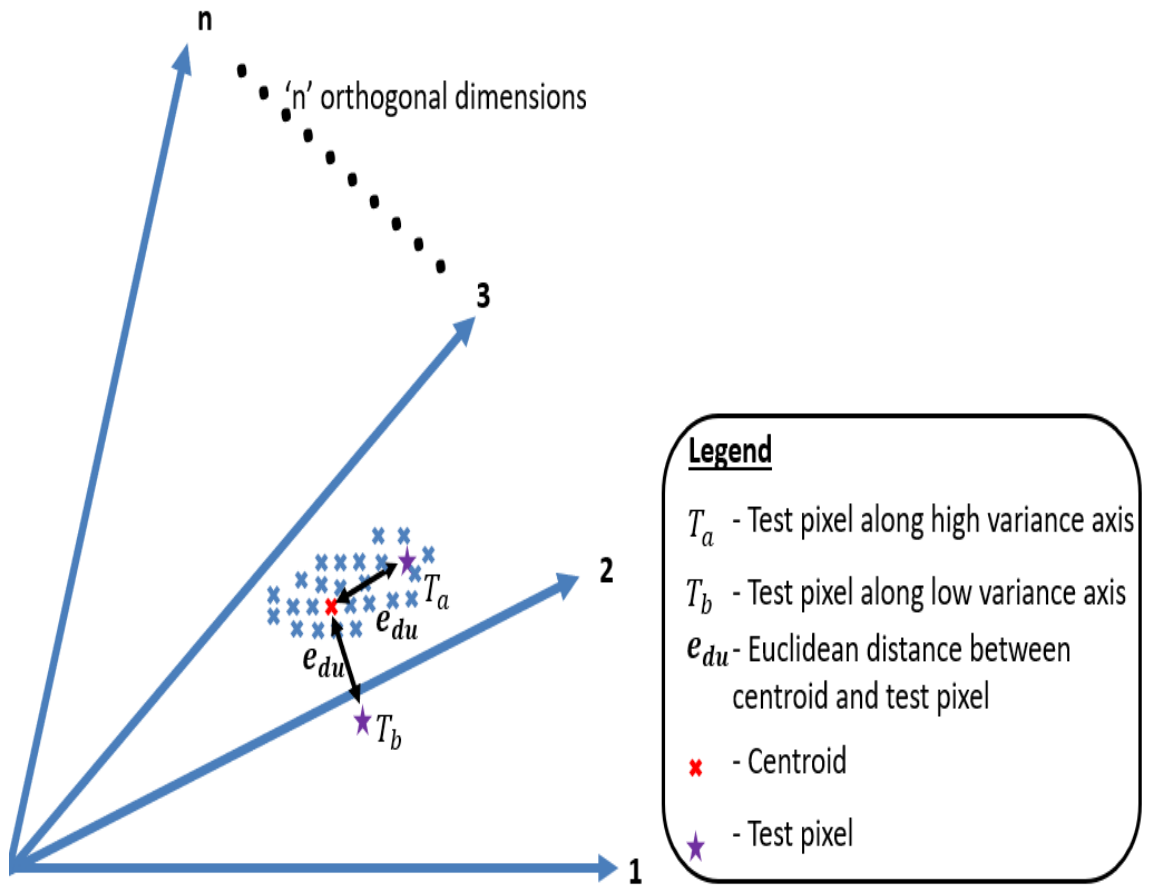




**Figure 3.1** Flowchart showing the implementation of Euclidean distance based method for classification of hyperspectral image.



**Figure 3.2** The Euclidean distance of a sample pixel from the centroid of the given balanced or symmetric distribution in an n-D space.



**Figure 3.3** The Euclidean distance of a sample pixel from the centroid of the given unbalanced or non-symmetric distribution in a n-D space.

### 3.1.3 Classification using Mahalanobis Distance

In such cases the Euclidean distance from the centroid along each dimension needs to be scaled by a factor to compensate for the variance along that dimension. Such type of variance compensated distance is called the Mahalanobis distance. Mahalanobis distance provides a scalar measure of the location of a multi-dimensional vector in a multi-variate parameter space. The advantages of using Mahalanobis distance for correlated data sets has already been demonstrated by Srivastava and Rao [61].

Mahalanobis distance was used for target detection from remotely sensed hyperspectral images by Meidunas [29] and to classify remotely sensed images by extracting information content in statistics of target and the background by Imani [30]. The utilisation of Mahalanobis distance to achieve classification of neutrophils from a hyperspectral medical image has been explored in the next paragraph.

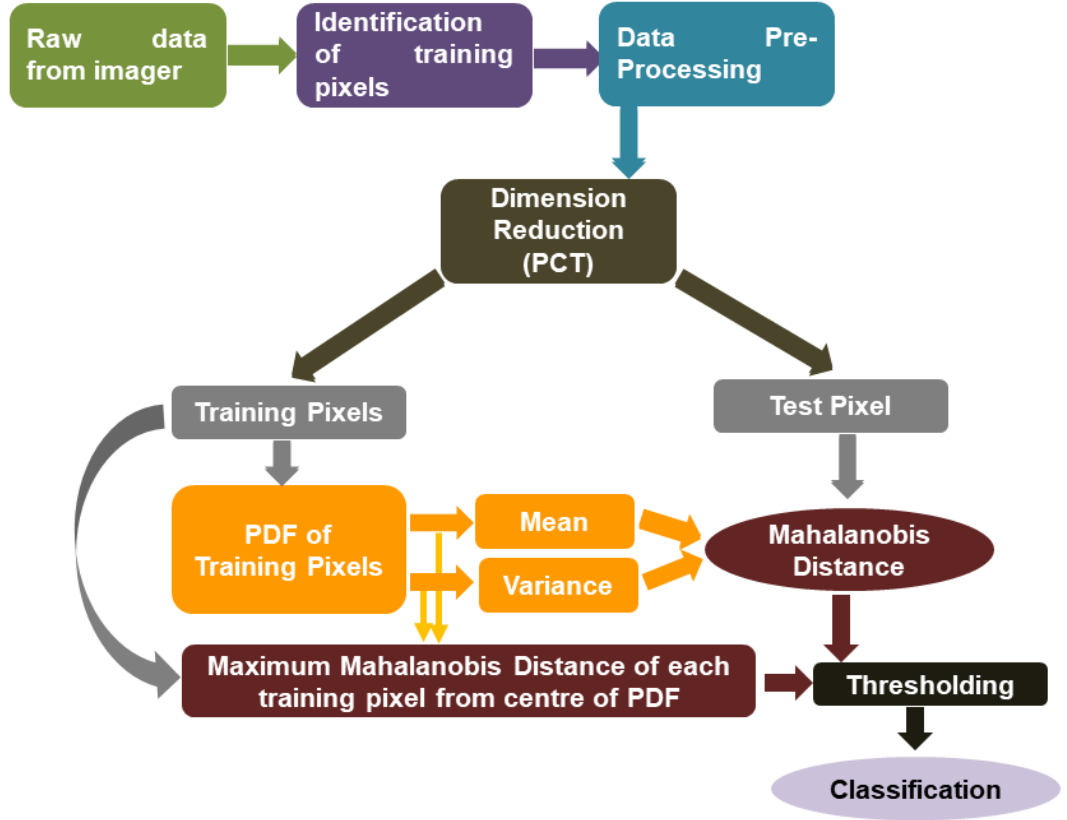
### 3.1.4 Implementation of Mahalanobis Distance Method for Classification

If the mean of the distribution of the CML-affected cytoplasm and nucleus is  $\mu_c$  and  $\mu_n$  respectively, and the covariance matrix of the CML-affected cytoplasm and the nucleus is  $\Sigma_c$  and  $\Sigma_n$  respectively, then the Mahalanobis distance of the test pixel,  $test(i, j)$  from the distributions of the CML-affected cytoplasm and nucleus was computed to be  $d_{mc}$  and  $d_{mn}$  as given in equations (3.4) and (3.5)[29,30].

$$d_{mc} = \sqrt{(test(i, j) - \mu_c)^T \Sigma_c^{-1} (test(i, j) - \mu_c)} \quad (3.4)$$

$$d_{mn} = \sqrt{(test(i, j) - \mu_n)^T \Sigma_n^{-1} (test(i, j) - \mu_n)} \quad (3.5)$$

A predefined threshold value,  $mf$  was multiplied with minimum value of Mahalanobis distance and cases of positive detection were highlighted based on equation



**Figure 3.4** Flowchart showing the implementation of Mahalanobis distance based method for classification of hyperspectral image.

(3.6). Decision on classification was taken based on a visual appreciation of the highlighted pixels.

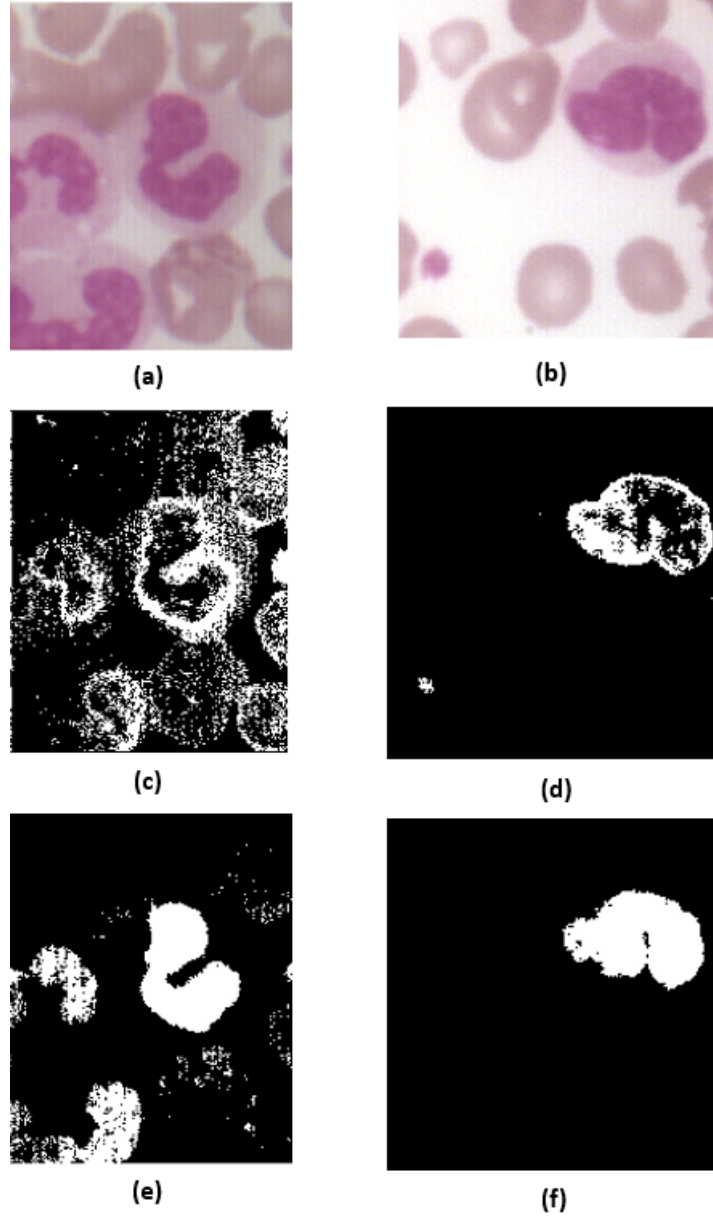
$$\left[ d_{mc} < mf \times \min(d_{mc}) \right] OR \left[ d_{mn} < mf \times \min(d_{mn}) \right] \implies Positive\ Detection \quad (3.6)$$

A schematic flow chart showing the sequence of steps required to be executed for implementing Mahalanobis distance method based classification algorithm on hyperspectral dataset for the purpose of classification is given at Fig. 3.4.

## 3.2 Results

The measures of sensitivity, specificity, accuracy and likelihood ratios were applied on the results compiled after application of the Euclidean and Mahalanobis distance methods on the 57 hyperspectral cube samples (obtained by 57 fold cross validation technique as given in section 1.4.3). Uniform classification parameters like the threshold multiplication factors and the size of the test samples were maintained during the classification procedure. All the cubes were reduced to seven high information bands by applying PCA. The results obtained by applying the Euclidean and the Mahalanobis distance methods on two sample hyperspectral sub-cubes is shown at Fig. 3.3. Fig. 3.5a, 3.5c and 3.5e shows the RGB image, the detection using Euclidean distance method and detection using Mahalanobis distance method in respect of the third hyperspectral sub-cube of the first hyperspectral cube sample. Fig. 3.5b, 3.5d and 3.5f shows the RGB image, the detection using Euclidean distance method and detection using Mahalanobis distance method in respect of the third hyperspectral sub-cube of the second hyperspectral cube sample.

The number of TP, TN, FP, FN cases found and the sensitivity, specificity and accuracy values derived from them after carrying out classification of the hyperspectral cubes using Euclidean and Mahalanobis distance methods have been tabulated (Table 3.1). A high LR (+) value indicates a significantly high diagnostic weight in arguing towards the presence of the disease (i.e., positive test for CML), with respect to Mahalanobis distance method as compared to the Euclidean distance method. On the contrary, a very low negative likelihood ratio [LR(-)] makes the Euclidean distance procedure better at ruling out the presence of disease (i.e., negative test for CML) as compared to Mahalanobis distance method (Table 3.2).



**Figure 3.5** Detection of CML neutrophil in two samples. (a) The RGB depiction of the third hyperspectral sub-cube of the first hyperspectral cube sample; (b) The RGB depiction of the third hyperspectral sub-cube of the second hyperspectral cube sample; (c) The detection of CML neutrophils in the third hyperspectral sub-cube of the first hyperspectral cube sample using Euclidean distance method; (d) The detection of CML neutrophils in the third hyperspectral sub-cube of the second hyperspectral cube sample using Euclidean distance method; (e) The detection of CML neutrophils in the third hyperspectral sub-cube of the first hyperspectral cube sample using Mahalanobis distance method; (f) The detection of CML neutrophils in the third hyperspectral sub-cube of the second hyperspectral cube sample.

**Table 3.1** The classification of hyperspectral data of the CML blood sample using Euclidean and Mahalanobis distances.

Distance Method	Classification				Sensitivity <sup>e</sup> (%)	Specificity <sup>f</sup> (%)	Accuracy <sup>g</sup>
	TP <sup>a</sup>	TN <sup>b</sup>	FP <sup>c</sup>	FN <sup>d</sup> (%)			
Euclidean	9	32	14	2	81.81	69.56	71.93
Mahalanobis	6	41	5	5	54.54	89.13	82.45

<sup>a</sup> TP: True Positive

<sup>b</sup> TN: True Negative

<sup>c</sup> FP: False Positive

<sup>d</sup> FN: False Negative

<sup>e</sup> Sensitivity =  $TP/(TP+FN)$

<sup>f</sup> Specificity =  $TN/(TN+FP)$

<sup>g</sup> Accuracy =  $(TP+TN)/(TP+TN+FP+FN)$

**Table 3.2** Likelihood ratios of classification using Euclidean and Mahalanobis distance methods.

Distance Method	LR(+) <sup>a</sup>	LR(-) <sup>b</sup>
Euclidean	2.68	0.26
Mahalanobis	5.02	0.51

<sup>a</sup> LR(+): Positive Likelihood Ratio  
= Sensitivity/(1-Specificity)

<sup>b</sup> LR(-): Negative Likelihood Ratio  
= (1-Sensitivity)/Specificity



### 3.3 Discussion

Euclidean and Mahalanobis distances have been extensively used in clustering, pattern recognition and classification problems [29-30]. They have been used for biomedical applications like medical image analysis to process gray scale, RGB and multispectral images. However, their application for analysis and classification of hyperspectral images is mostly restricted to remote sensing applications like land use classification, mineralogy and target detection [30-32]. Remotely sensed hyperspectral images have low spatial resolution and can be used to classify larger objects like plantation or crop coverage, mineral deposits and military target areas [19,23,30-32]. Biomedical hyperspectral images are obtained by mounting the hyperspectral sensor on a microscope. This generates hyperspectral image cubes having high spectral as well as high spatial resolution. Such image cubes are in the form of huge data sets. Also the objects being classified are cells and tissues at the micron level, whereas, much larger objects are classified in remotely sensed hyperspectral images. Therefore the processing of the hyperspectral image cubes of biological tissues and cells require robust algorithms and are time consuming. The use of statistical distances for classifying high spatial and spectral resolution hyperspectral biomedical images has been introduced in this thesis.

Statistical distances exploit the morphology of the image cube to derive classification metrics. The use of spectral nuclear morphometry to classify lymphocytes in lymphoproliferative disorders (i.e., chronic lymphatic leukemia, infectious mononucleosis and non-Hodgkin's lymphoma) was studied by Greenspan et al. [35]. Their work focused only on the nuclear region of the cells and showed that computational analysis of spectral data brought out the spectral and morphological patterns associated with each lymphoproliferative disorder. Malik et al. studied chronic lymphocytic leukemia (B-CLL) and normal small lymphocytes by analyzing their morphological patterns [36]. These two studies focused only on the nucleus of lymphocytes whereas the present study analyzed spectral data from both the nucleus

and cytoplasm of neutrophils. Though analyzing the nucleus and cytoplasm may have added some ‘noise’ to the data, the distance methods have shown sufficient promise to consider the gathering of hyperspectral image data from a larger pool of samples.

### **3.3.1 Comparison with Windowed SAM and 3-D SGM Methods**

The higher sensitivity values of the classification methods based on statistical distances indicates better ability to detect positive CML cases as compared to the the Windowed SAM and the 3-D SGM methods. Also, the low negative likelihood ratios of both the statistical distance methods, makes them ideal for ruling out the presence of CML in a sample, as compared to the Windowed SAM and the 3-D SGM methods. But, with lower specificity and accuracy values, the Euclidean Distance method is less robust in detecting healthy samples and segregating the CML affected samples from the healthy samples as compared to the Windowed SAM and the 3-D SGM methods. However, Mahalanobis distance method is better than the Windowed SAM method in detecting healthy samples and segregating healthy from diseased samples due to its higher specificity and accuracy values.

### **3.3.2 Comparison between Euclidean and Mahalanobis Distance Methods**

In this study, the accuracy and specificity of Mahalanobis distance method was found to be higher than that of the Euclidean distance method. This indicates a greater ability to detect healthy neutrophils and clearly classify neutrophils of CML patients from that of healthy patients. This is because Mahalanobis distance method is based on classifying the test pixel based on its spectral similarity to a distribution rather

than to a single pixel. It is best suited for probabilistic classification where the probability distribution has different variances along different dimensions. It is also ideal for initial diagnostics to distinguish healthy patients from those suffering from CML. The sensitivity of Euclidean distance method is greater than Mahalanobis distance method, and makes it best suited for confirmatory diagnostics to detect positive cases.



## Chapter 4

# Classification of CML Neutrophils using Frequency Domain Normal Probability Distribution Function Merged with Mahalanobis Distance Method

### 4.1 Introduction

The classification algorithms explored in the above chapters were in the spatial domain. Of all classification algorithms in spatial domain that were discussed in the previous chapters, the Mahalanobis distance method was found to have a very low computation time. It was found to have a very low specificity (89.13%) and accuracy (82.4%). However it had a very low sensitivity measure of 54.54%. To improve the sensitivity measure, it was necessary to exploit the information content in the frequency domain. Therefore, the Frequency Domain Normal Probability Distribution Method (FDNPM) metric was proposed for classification. However, FDNPM needed to be combined with Mahalanobis distance method for classification

to leverage the higher specificity and accuracy values. Therefore, a merged method using the FDNPM metric and Mahalanobis distance was proposed for classification of CML neutrophils.

## **4.2 Use of Mahalanobis Distance Method in Classifying Multi-dimensional Images and its Limitations**

The classification of CML neutrophils using Mahalanobis distance on hyperspectral data sets has already been explored in the spatial domain by Panda et. al. [45]. However, the method suffered from low sensitivity values which indicated reduced capability for segregating CML positive cases. Mahalanobis distance is an effective measure of how close a sample is to the Probability Density Function (PDF) of the CML affected neutrophils. However it was used in the spatial domain and did not exploit the information content in the frequency domain. Also the Mahalanobis distance method calculates the distance of a single pixel from a given PDF [29-30]. The CML affected neutrophils are actually a set of pixels. Therefore it was necessary to compute the combined similarity of a set of pixels with the PDF of the CML affected neutrophils.

## **4.3 Probability distribution function (PDF)**

PDF is a mathematical function that gives the probability of occurrence of any event. In the present study the probability distribution function of the Fourier transform of the sample pixels gives the probability that a given frequency domain response obtained from the Fourier transform of the test sample would have the intensity as defined by the argument of the PDF. Since the groups of pixels being considered

were in each others neighbourhood, their spectral characteristics were found to be similar. Therefore, their probability distribution was found to be normal or near normal centred around a mean spectral intensity with varying standard deviations. Hence, the sample groups of pixels being considered were modeled into a normal distribution with various means and standard deviations as applicable.

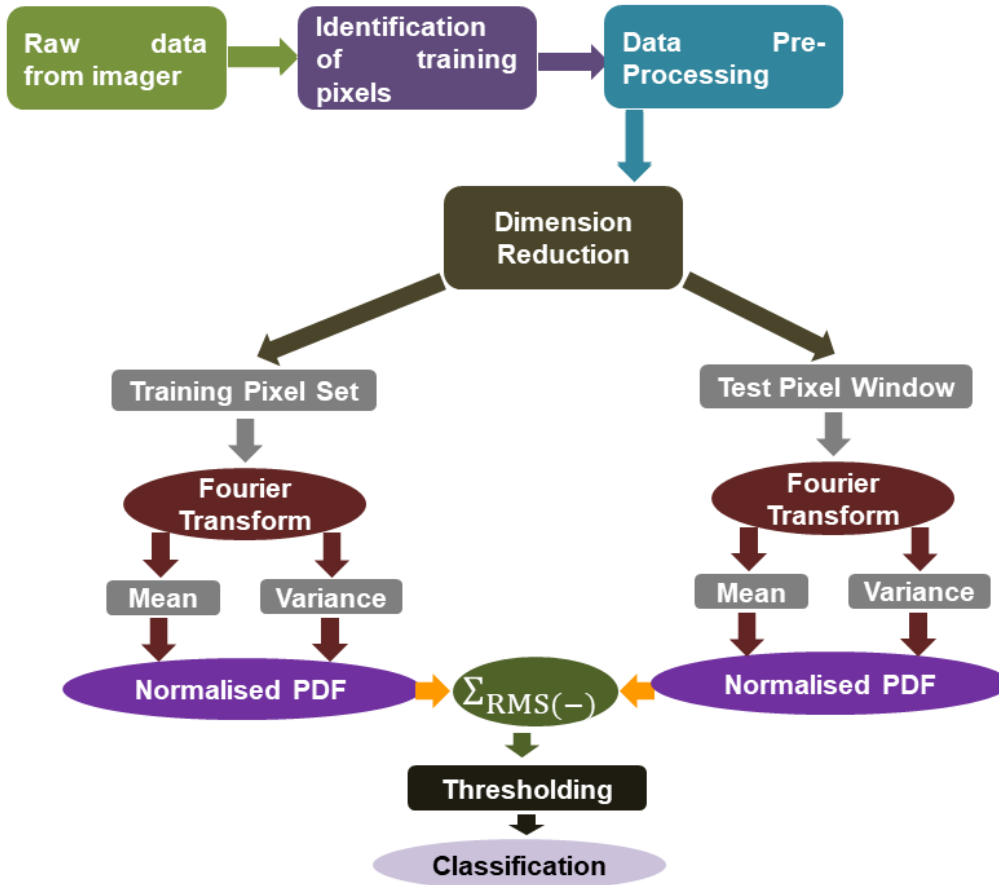
## 4.4 FDNPM

The first step of the FDNPM method was to match the probability distribution of the training and the test pixels in the frequency, along each of the reduced dimensions. To do the same, the Fourier transform of the training and the test samples was carried out before computing the normalised PDF. Fourier transform of the hyperspectral cube yields the original information content re-framed as sum of weighted frequency responses, i.e., the rates of change of spatial data. The implementation in frequency domain helped in exploiting the information content in the various rates of change of spatial intensity values of the hyperspectral image cube. This significantly improved the sensitivity values.

A schematic flow chart showing the sequence of steps required to be executed for implementing FDNPM method on hyperspectral dataset for the purpose of classification is given at Fig. 4.1.

## 4.5 Merged Method

The Mahalanobis distance is the most time efficient algorithm for classifying hyperspectral images. However, when used for classification, it had a very low sensitivity. Therefore, to improve the sensitivity values, the present study explores detection using an algorithm to bring out spectral similarity measure of a pixel to the spectral domain PDF of the CML neutrophils, merged with the spatial domain Mahalanobis



**Figure 4.1** Flowchart showing the implementation of FDNPM method for classification of hyperspectral image.



distance method. FDNPM was found to be an ideal method to bring out spectral similarity measure in frequency domain. Therefore, to improve the efficacy of the Mahalanobis distance method and to exploit the information content in frequency domain, the use of PDF matching of the Fourier transform of test and the training hyperspectral samples along with the Mahalanobis distance in spatial domain has been proposed in this chapter. A combination of FDNPM and Mahalanobis distance method called the merged method, has been proposed in this chapter.

## 4.6 Implementation of Merged Method

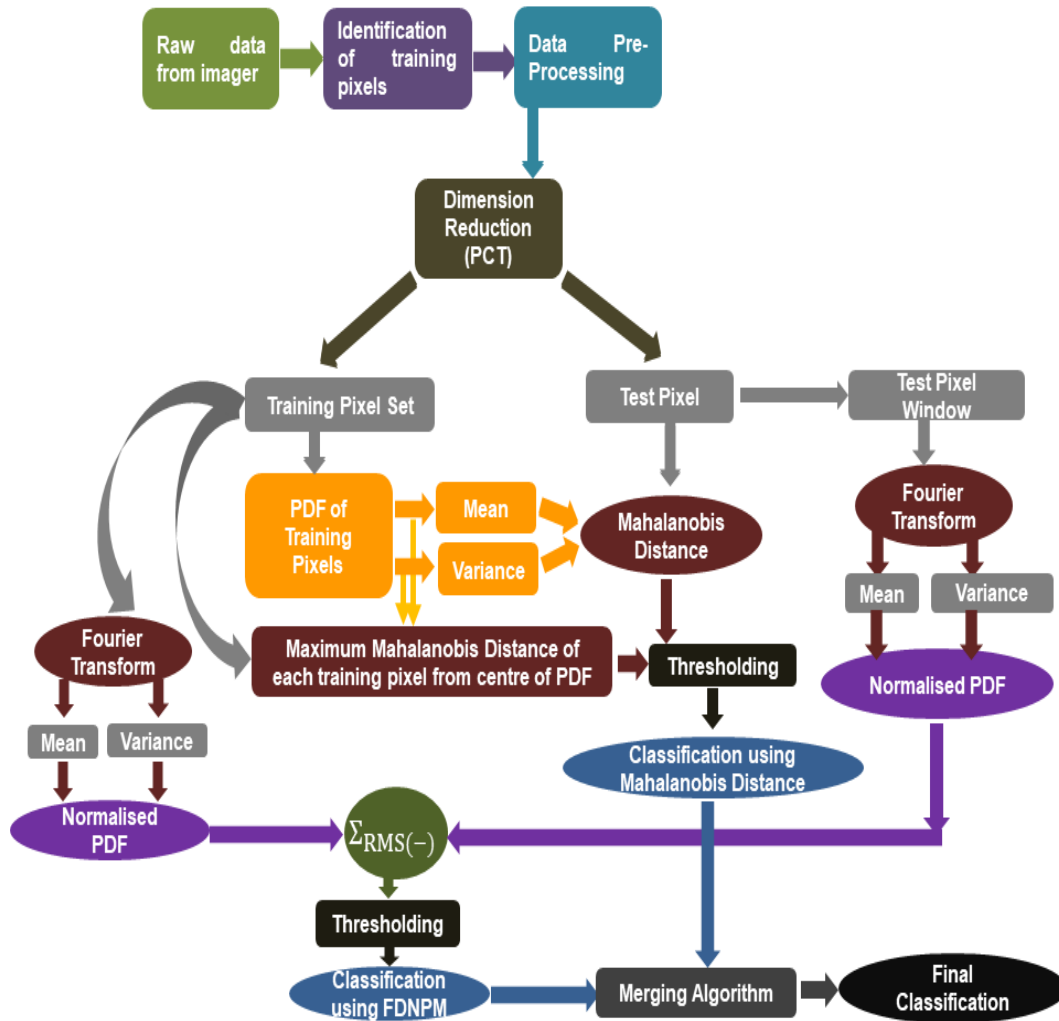
### 4.6.1 PDF of training pixels

Four sets of training pixels were identified for the neutrophil and the cytoplasm. Their Fourier transform was computed. The mean and standard deviations of Fourier transform of these sets of training pixels were computed in each band and used to generate the band wise normal PDF for both cytoplasm and neutrophil over 2047 intensity values as shown at equation (4.1), (4.2) and (4.3).

A schematic flow chart showing the sequence of steps required to be executed for implementing Merged method based on FDNPM and Mahalanobis distance methods for classification of hyperspectral dataset for the purpose of classification is given at Fig. 4.2.

$$MTtraining_{i,j} = mean(FT(training\ pixel\ set_{i,j})) \quad (4.1)$$

$$STtraining_{i,j} = standard\ deviation(FT(training\ pixel\ set_{i,j})) \quad (4.2)$$



**Figure 4.2** Flowchart showing the implementation of Merged method comprising of the Mahalanobis distance method and the FDNPM method for classification of hyperspectral image.

$$NDtraining_{i,j} = normal\ PDF(MTtraining, STtraining, -1023\ to\ 1023) \quad (4.3)$$

where,

$MTtraining_{i,j}$ : The mean of the set of set of  $10 \times 11$  training pixels starting from the pixel  $(i, j)$  in reduced bands.

$FT(x)$ : The fourier transform of the pixels given by the variable,  $x$ .

$mean(\theta)$ : The mean of the set of values given by  $\theta$ .

$standarddeviation(\theta)$ : The standard deviation of the set of values given by  $\theta$ .

$STtraining_{i,j}$ : The standard deviation of the set of  $10 \times 11$  training pixels starting from the pixel  $(i, j)$  in reduced bands.

$NDtraining_{i,j}$ : The normal PDF of the Fourier transform of the set of  $10 \times 11$  set of training pixels starting from the pixel,  $(i, j)$  in reduced bands over 2047 intensity values from -1023 to 1023 having mean given by  $MTtraining_{i,j}$  and standard deviation given by  $STtraining_{i,j}$ .

#### 4.6.2 PDF of Test Pixels

A window of  $10 \times 11$  pixels were earmarked for each pixel at the location  $(i, j)$  of the test hyperspectral image in reduced dimensions keeping the pixel at  $(i, j)$  at top left corner of the window. The Fourier transform of the window of test pixels was computed. The mean and standard deviations of these window of test pixels sets

were computed in each band and used to generate the band wise normal PDF over 2047 intensity values as shown in equation (4.4), (4.5) and (4.6).

$$MTtest_{i,j} = mean(FT(test\ pixel\ set(i,j))) \quad (4.4)$$

$$STtest_{i,j} = standard\ deviation(FT(test\ pixel\ set(i,j))) \quad (4.5)$$

$$NDtest_{i,j} = normal\ PDF(MTtest, STtest, -1023\ to\ 1023) \quad (4.6)$$

where,

$MTtest_{i,j}$ : The mean of the set of set of  $10 \times 11$  test pixels starting from the pixel  $(i, j)$  in reduced bands.

$STtest_{i,j}$ : The standard deviation of the set of  $10 \times 11$  test pixels starting from the pixel  $(i, j)$  in reduced bands.

$NDtest_{i,j}$ : The normal PDF of the Fourier transform of the set of  $10 \times 11$  set of test pixels starting from the pixel,  $(i, j)$  in reduced bands over 2047 intensity values from -1023 to 1023 having mean given by  $MTtest_{i,j}$  and standard deviation given by  $STtest_{i,j}$ .

$normalPDF(\alpha, \beta, \gamma)$ : The normal PDF having mean, standard deviation and range of pixel intensity values given by  $\alpha, \beta$ , and  $\gamma$  respectively.

### 4.6.3 Matching frequency domain PDF of the training and test pixels

The PDF obtained from the frequency domain response of the window of training pixels was compared with the PDF of the frequency domain response of the sets of test pixels by taking the root mean square (RMS) difference. A threshold of the RMS difference was earmarked. If the RMS difference corresponding to any test pixel as given in equation (4.7), was found to be less than the threshold RMS difference, it was marked to be a positive detection for CML neutrophil.

$$RD(i, j) = RMS \left[ (ND_{test_{i,j}}) - (ND_{training_{i,j}}) \right] \quad (4.7)$$

where,

$RD(i, j)$ : The RMS difference of the PDF of the Fourier transform of the set of test and the window of training hyperspectral pixels across the seven PCT bands corresponding to the pixel at  $(i, j)$  position.

### 4.6.4 Merging of Mahalanobis and frequency domain PDF matching methods

In order to leverage the advantages of both the techniques, they were merged in a manner such that Mahalanobis distance method was given a weightage of 1.3 and normal FDNPM method was given a weightage of 1 as given in equation (4.8).

$$merged\ method = (mahal \times normpdf) + (1.3 \times mahal) + normpdf \quad (4.8)$$

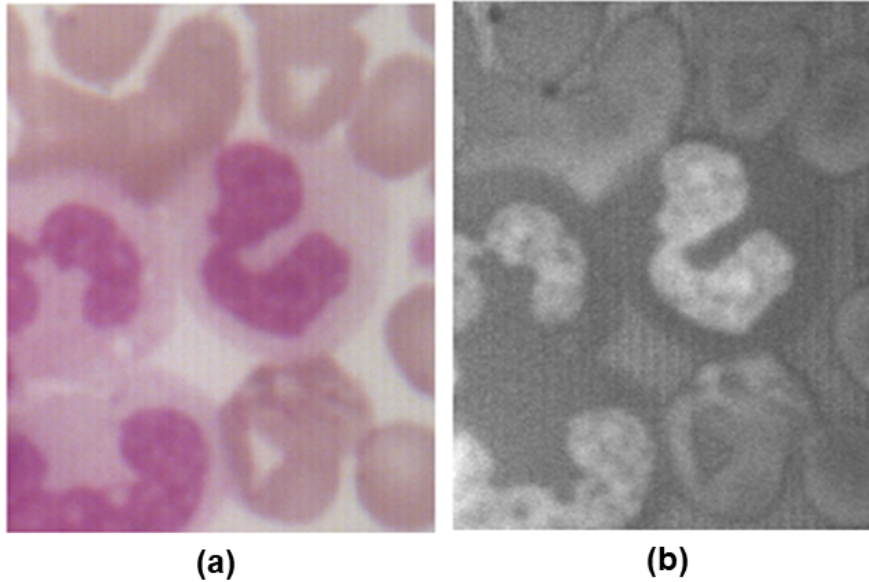
where,

*mergedmethod*: The results obtained using merged method comprising of Mahalanobis distance and FDNPM methods

*mahal*: The results obtained by using Mahalanobis distance method

*normpdf*: The results obtained by using FDNPM method.

The first term in the right hand side of equation (4.8) ensures that pixels of zero intensity in any of the results remain same in the final result while the balance of the intensity values are appropriately scaled. The second and the third terms in equation (4.8) give a weightage of 1.3 to the Mahalanobis distance method and a weightage of 1 to the FDNPM method. The RGB image of the third hyperspectral sub cube of the first hyperspectral cube is shown at Fig. 4.3a. The detection results using the merged method is depicted in Fig. 4.3b respectively. For the merged method, a threshold value of 1.26 for the FDNPM and 1.20 for the Mahalanobis distance methods were fixed.



**Figure 4.3** The depiction of (a) RGB image of the third hyperspectral sub-cube obtained from the first hyperspectral cube and (b) the merged method (using Mahalanobis distance and FDNPM methods).

**Table 4.1** Sensitivity, specificity and accuracy of the KCED and the merged (Mahalanobis distance + FDNPM) methods on healthy versus CML neutrophils.

Method	Classification parameters						
	TP <sup>a</sup>	FP <sup>b</sup>	TN <sup>c</sup>	FN <sup>d</sup>	Sensitivity <sup>e</sup> (%)	Specificity <sup>f</sup> (%)	Accuracy <sup>g</sup> (%)
<b>Mahalanobis Distance</b>	6	14	32	2	63.6	93.5	87.7
<b>Mahalanobis +FDNPM*</b>	7	3	43	4	63.6	93.5	87.7

<sup>a</sup> TP: True positive

<sup>b</sup> FP: False positive

<sup>c</sup> TN: True negative

<sup>d</sup> FN: False negative

<sup>e</sup> Sensitivity =  $TP/(FN+TP)$

<sup>f</sup> Specificity =  $TN/(FP+TN)$

<sup>g</sup> Accuracy =  $(TN+TP)/(TP+FP+TN+FN)$

\* FDNPM: Frequency Domain Normal PDF (Probability Distribution Function) Matching.

## 4.7 Results

### 4.7.1 Sensitivity, specificity and accuracy

The merged method significantly improved the sensitivity of Mahalanobis distance method from 54.5% to 63.6% without any significant reduction in the specificity (93.5%) and accuracy values (87.7%). Therefore the merged method significantly improved the probability of positive diagnosis of the sample as given in Table 4.1.

### 4.7.2 Likelihood ratios

The merged method had a lower negative likelihood ratio LR(-) as compared to Mahalanobis distance method resulting in an improved ability to defend a negative test for CML (i.e., rule out the presence of the disease). However, it displayed a lower positive likelihood ratio LR(+) indicating a lower diagnostic weight in defending a positive test for CML (Table 4.2).

**Table 4.2** Comparison of likelihood ratios of the Mahalanobis distance method and the merged (Mahalanobis distance + FDNPM) methods.

Method	LR(+) <sup>a</sup>	LR(-) <sup>b</sup>
<b>Mahalanobis distance method</b>	5.02	0.51
<b>Mahalanobis + FDNPM*</b>	9.8	0.4

<sup>a</sup> LR(+): Positive Likelihood Ratio = Sensitivity/(1-Specificity)

<sup>b</sup> LR(-): Negative Likelihood Ratio = (1-Sensitivity)/Specificity

\* FDNPM: Frequency Domain Normal PDF (Probability Distribution Function) Matching.

## 4.8 Discussion

We had previously explored the use of single-classifier approaches (the Euclidean and Mahalanobis distances) to correctly identify and group CML from healthy neutrophils [45]. The Mahalanobis distance method had higher specificity (89%) and accuracy (82.5%) performance values compared to the results for the Euclidean distance algorithm of 69.6% and 71.9%, respectively. As far as sensitivity was concerned, the Euclidean distance had a 18.2% chance of missing a positive diagnosis versus 45.5% for the Mahalanobis distance. On the whole, neither method was stronger for all three performance parameters. Subsequently, a novel method (3-D Spectral Gradient Mapping) was introduced in chapter 3 [45], increased the specificity (97.7%) and accuracy (84.2%) of classification but dramatically dropped when it came to its sensitivity. In the current work, a few corrective measures were introduced and has shown to produce stronger classification percentages by decreasing false positive and negative values. Since the same hyperspectral data sets were used, a direct comparison could be made across the different algorithms.

In the current study, both the Mahalanobis distance and the FDNPM methods use PDF for deriving the classification metrics. The Mahalanobis distance method computes the similarity index by calculating the Mahalanobis distance of a test pixel from the PDF of the training pixel to classify it. However, the whole neutrophil cell per se is represented by a set of pixels. To classify them correctly, it is necessary to



compute its similarity with the set of training pixels. The FDNPM method computes the similarity by considering the entire set of test pixels for the nucleus and the cytoplasm of the neutrophils. Further, the Fourier transform of the hyperspectral cube represents the information content as sum of scaled frequency responses. The frequencies give the rate of change of information over the space, i.e., along  $X$  (the rows),  $Y$  (the columns) and  $Z$  (the wavelength bands) axes. Therefore, by matching the PDF in the frequency domain, the spatial rate of change of the information content of the hyperspectral cube was compared. The Mahalanobis distance has been merged with the FDNPM method to improve on the sensitivity value without compromising the former's high specificity and accuracy values. In fact, the merged method was found to have increased the specificity (by 4.4%) and accuracy (by 5.2%) values as well. The merged method has much higher LR(+) values (9.80) than solely using the Mahalanobis distance approach (5.02). With a lower LR(-) value (0.40), FDNPM method improvises the Mahalanobis distance method by reducing the likelihood of missing diseased cases (compared to only using the Mahalanobis distance;  $\text{LR}(-) = 0.51$ ; [48]).



## Chapter 5

# Classification of CML Neutrophils using Kurtosis Compensated Euclidean Distance

### 5.1 Introduction

Use of euclidean distance over multiple dimensions as the measurement metric for classification of hyperspectral cubes for CML detection has been discussed in chapter 2. In such cases the Euclidean distance of the test pixel is measured from the mean of the training pixel sets and compared with a constant threshold distance. However, distributions of the training pixel sets around its mean is not always constant. Therefore, mean of a distribution is not a single point identification for the entire distribution. Thus, the use of euclidean distance computed from the mean of the distribution for classification would not give the correct results. The shape of the distribution should decide the value the threshold distance. In other words keeping the threshold distance constant, the measured distance needs to be suitably scaled based on the shape of the PDF and compared with the constant threshold distance for classification.

### 5.1.1 Classification using KCED

KCED is a metric derived from the probability distribution which calculates the Euclidean distance of any sample from a given probability distribution by compensating for the spread of the distribution along any given direction. A similar measure of classification is Diffusion Kurtosis Imaging and Diffusion Tensor Imaging. These classification methods were used to categorize biological tissues based on non-Gaussian diffusion of water in biological systems [54-56].

#### Euclidean distance

Euclidean distance gives the shortest distance between two points in multi-dimensional space. The Euclidean distance,  $ed_{test,trg}$  between the test pixel,  $test(i, j)$  and a training pixel,  $trg(i, j)$  in  $n - D$  space is given by equations (5.1) and (5.2) [45].

$$ed_{test,trg}(i, j) = \sqrt{\sum_{k=1}^n (test(i, j, k) - trg(i, j, k))^2} \quad (5.1)$$

where,

$ed_{test,trg}(i, j)$ : Euclidean distance between the test pixel and the training pixel.

$test(i, j, k)$ : The intensity value of the test pixel at the coordinates  $(i, j)$  in the  $k^{th}$  wavelength band.

$trg(i, j, k)$ : The intensity value of the training pixel at the coordinates  $(i, j)$  in the  $k^{th}$  wavelength band.

$$\implies ed_{test,trg}(i, j) = \sqrt{(test(i, j) - trg(i, j))^T (test(i, j) - trg(i, j))} \quad (5.2)$$

where,

$test(i, j)$ : The intensity value of the test pixel at coordinates  $(i, j)$  in all the wavelength bands.

$trg(i, j)$ : The intensity value of the training pixel at coordinates  $(i, j)$  in all the

wavelength bands.

For calculating the Euclidean distance between test pixel,  $test(i, j)$  and a set of training pixels, the distance was calculated between  $test(i, j)$  and a single pixel representing the set of training pixels based on equation (5.2) across all wavelength bands. This single pixel can be safely assumed to be the mean,  $\mu_{trg}$  of the probability distribution formed by the set of training pixels. Thus the Euclidean distance of the test pixel,  $test(i, j)$  from the probability distribution of training pixels having mean  $\mu_{trg}$  is given by equation (5.3).

$$epdf_{test, trg}(i, j) = \sqrt{(test(i, j) - \mu_{trg})^T (test(i, j) - \mu_{trg})} \quad (5.3)$$

where,

$\mu_{trg}$ : Mean of the set of training pixels.

$epdf_{test, trg}$ : Euclidean distance of the test pixel,  $test(i, j)$ , from the mean of the PDF of the training pixels,  $\mu_{trg}$ , across all wavelength bands.

### ***Kurtosis***

Kurtosis of a probability distribution is a measure of the heaviness of the tail of the probability distribution [54-56]. The Kurtosis of a probability distribution,  $X$  is the scaled fourth central moment and is given by equation (5.4).

$$K(X) = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] \quad (5.4)$$

where,

$\mu$ : The mean of the probability distribution  $X$ .

$\sigma$ : The variance of the probability distribution  $X$ .

For a PDF with zero mean the Kurtosis can be defined as given in equation (5.5).

$$K(X) = \frac{\mu^4}{\sigma^4} \quad (5.5)$$

### ***KCED***

The Kurtosis of a normal distribution where the tail of the distribution is equally or symmetrically distributed along all the dimensions is 3. Any reduction in tail of a distribution as compared to the normal distribution is reflected by a Kurtosis value less than 3. Such a distribution is classified as a platykurtic distribution. Any increase or heaviness in tail of a distribution as compared to the normal distribution is reflected by a Kurtosis value greater than 3. Such a distribution is classified as a leptokurtic distribution. Therefore, the actual distance of a sample point from a distribution can be obtained by dividing the Euclidean distance measure by the deviation in Kurtosis value from 3 along each dimension. Such a deviation is called excess Kurtosis. However, in the present study a variation of excess Kurtosis is obtained by dividing the Kurtosis value along each dimension by 3. Thus the Euclidean distance gets eclipsed when the tail is heavy and magnified when the tail is reduced as compared to a normal distribution. Therefore, the KCED,  $kepdf_{test,trg}$  of a point from the probability distribution of the training pixels in multi-dimensional space is given by equation (5.6).

$$kepdf_{test,trg} = \left[ \sum_{k=1}^M \left( \frac{(test(i,j,k) - trg(i,j,k))^2}{K_k(X)} \right) \right]^{1/2} \quad (5.6)$$

where,

$K_k(X)$ : The excess Kurtosis of the PDF of set of hyperspectral pixels in  $k^{th}$  dimension or band.

$M$ : Total number of dimensions or bands.

An excess Kurtosis matrix which has as many columns as the number of hyperspectral image pixels and each column of the matrix consisting of inverse of excess Kurtosis in each dimension as given at equation (5.7) was defined.

$$K(X)_{inv} = \begin{bmatrix} 3/K_1 & 3/K_1 & . & . & .Ntimes & . & . & 3/K_1 \\ 3/K_2 & 3/K_2 & . & . & .Ntimes & . & . & 3/K_2 \\ . & . & . & . & . & . & . & . \\ 3/K_M & 3/K_M & . & . & .Ntimes & . & . & 3/K_M \end{bmatrix} \quad (5.7)$$

where,

$N$ : Number of hyperspectral image pixels.

Therefore, KCED of a point from the probability distribution of the training pixels in multi-dimensional space can also be shown as given by equation (5.8).

$$kepdf_{test,trg}(i, j) = \sqrt{(test(i, j) - \mu_{trg})^T (K(X)_{inv}) (test(i, j) - \mu_{trg})} \quad (5.8)$$

In the present study, the KCED was calculated for all the pixels of the sample hyperspectral image from the probability distribution of training pixels for the cytoplasm and the nucleus. The pixels, spectrally similar to the training pixels, were found to have smaller distances as compared to the spectrally dissimilar pixels. The composite distance image was computed by adding the distance images for the cytoplasm and the nucleus in the first three PCT bands. The sum of the composite distance images in all the three PCT bands was calculated and plotted as the resultant distance image. Classification of a blood sample was made after visually appreciating the resultant distance image.

### 5.1.2 Scaling of Euclidean distance using Kurtosis

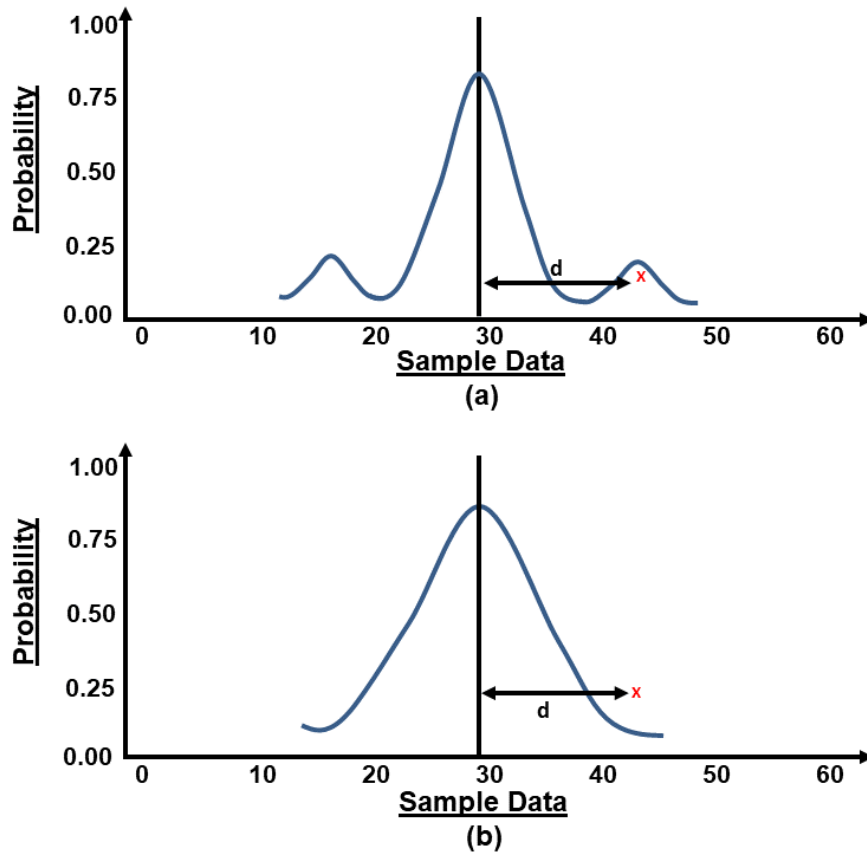
Based on the PDF analysis of random hyperspectral image samples, we found that the data had either normal or near normal distributions. The threshold for classification using Euclidean distance for normal distributions varied based on the standard deviation of the PDF. A higher Kurtosis along a given dimension indicated a heavy-tailed PDF along that dimension and vice versa. This meant that a higher threshold value for detection was required using the Euclidean distance along that dimension. The use of the same Euclidean distance with different Kurtosis would result in an error in classification (Fig. 5.1). Therefore, by scaling down the Euclidean distance for heavy-tailed distribution and scaling it up along a given dimension for a light tailed distribution, refined the classification procedure. Since Kurtosis represented the nature of tail of a normal PDF, compensation of the Euclidean distance using Kurtosis along each dimension aided in correcting the error.

The Euclidean distance gave the correct proximity measure of the test pixel from the probability distribution of the training pixel set, only if the mean of the distribution represents the actual centre of mass of the training pixel set. This was the case when the distribution was equally or symmetrically distributed along all dimensions about its mean. If not, the mean did not reflect the actual centre of distribution mass. In such cases also, it was necessary to compensate for the distance along each dimension by the excess distribution of pixels along that dimension.

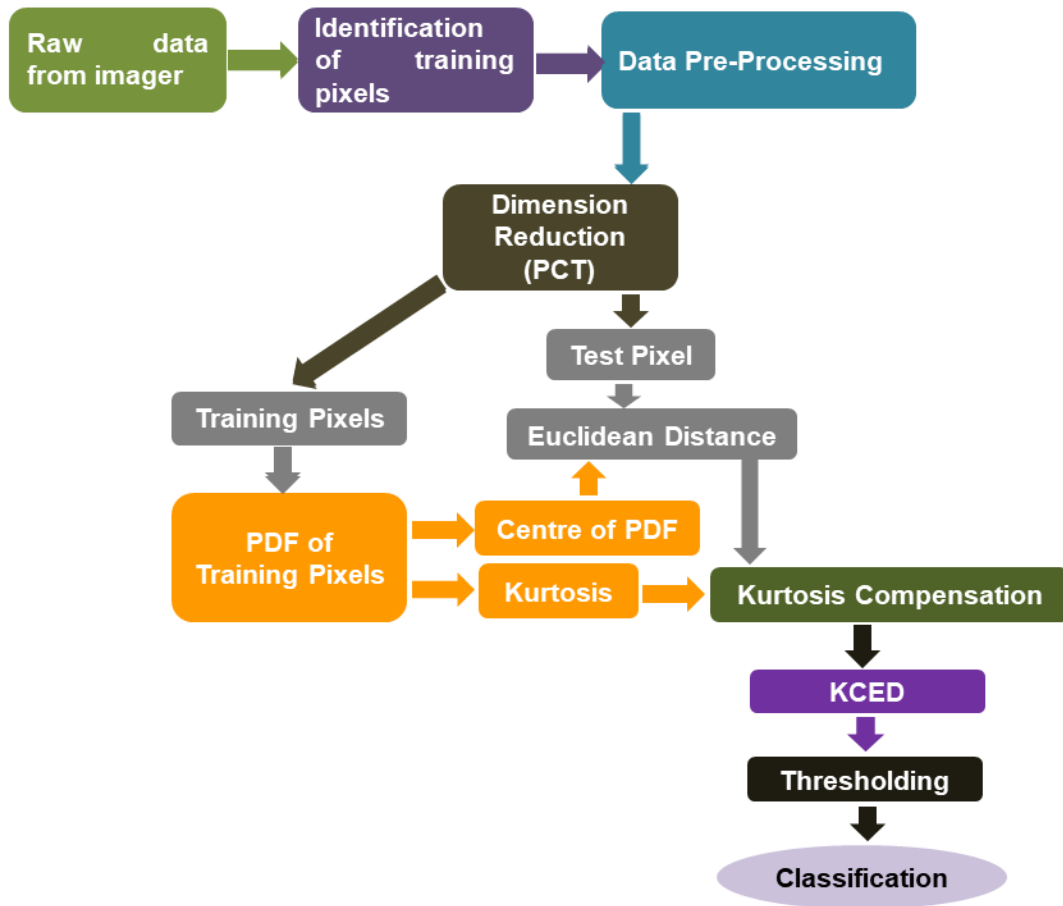
A schematic flow chart showing the sequence of steps required to be executed for implementing KCED method on hyperspectral dataset for the purpose of classification is given at Fig. 5.2.

The PDF of one set of training pixels for the cytoplasm and nucleus of CML neutrophils over the seven PCT bands are shown in Figs. 5.3a to g.

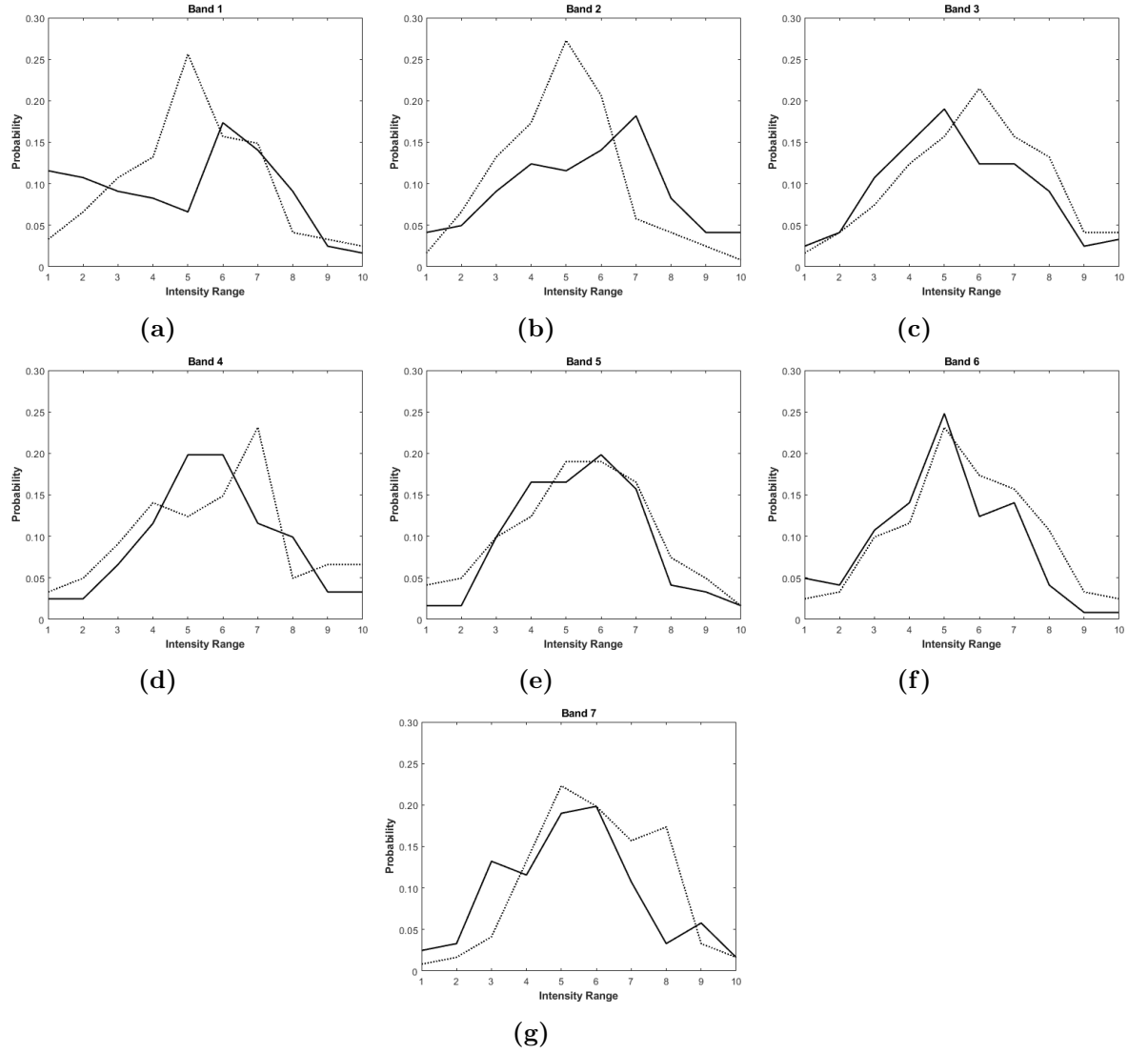




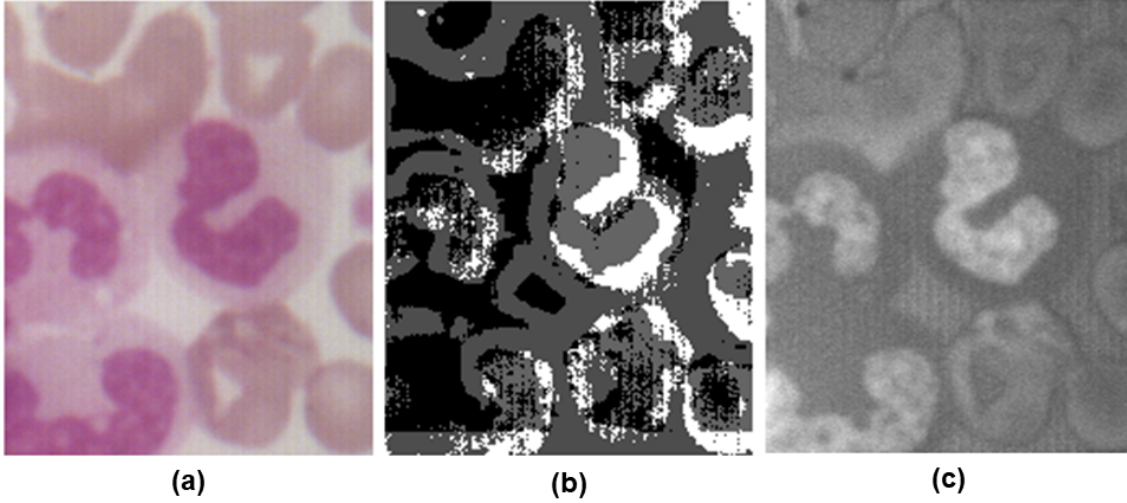
**Figure 5.1** The sample  $x$  is part of the heavy-tailed PDF in (a) but not a part of the light-tailed PDF in (b). The use of same Euclidean distance for PDF with different Kurtosis may result in incorrect classification results [ $d$  = distance of  $x$  from centre of the normal PDF]



**Figure 5.2** Flowchart showing the implementation of KCED method for classification of hyperspectral image by applying kurtosis compensation on Euclidean distance.



**Figure 5.3** The PDF across seven PCT bands [(a) to (g)] of one set (as a representation) of training pixels for cytoplasm (dotted line) and nucleus (solid line) of CML neutrophils.



**Figure 5.4** The depiction of (a) RGB image of the third hyperspectral sub-cube obtained from the first hyperspectral cube, (b) the merged method (using Mahalanobis distance and FDNPM methods) and (c) the detection of CML neutrophils using KCED.

### 5.1.3 Uniformity of parameters

Uniform parameters, like the threshold multiplication factors, reduction of dimension to seven bands by applying PCA and standardization of size of test samples to  $10 \times 11 \times 7$ , was ensured while applying the classification algorithms. The detection results obtained by using the merged method and KCED metric for classifying the third hyperspectral sub cube of the first hyperspectral cube were compared. The RGB image of the third hyperspectral sub cube of the first hyperspectral cube is shown at Fig. 5.4 (a). The detection results using the merged method and KCED are depicted in Fig. 5.4 (b) and (c) respectively. For the merged method, a threshold value of 1.26 for the FDNPM and 1.20 for the Mahalanobis distance methods were fixed.

### 5.1.4 Sensitivity, specificity and accuracy

It shared the second highest sensitivity value with euclidean distance method at 81.8% and the highest accuracy value of 91.2%. KCED was found to have the

**Table 5.1** Sensitivity, specificity and accuracy of the KCED method on healthy versus CML neutrophils.

Method	Classification parameters						
	TP <sup>a</sup>	FP <sup>b</sup>	TN <sup>c</sup>	FN <sup>d</sup>	Sensitivity <sup>e</sup> (%)	Specificity <sup>f</sup> (%)	Accuracy <sup>g</sup> (%)
<b>KCED<sup>*</sup></b>	9	3	43	2	81.8	93.5	91.2

<sup>a</sup> TP: True positive

<sup>b</sup> FP: False positive

<sup>c</sup> TN: True negative

<sup>d</sup> FN: False negative

<sup>e</sup> Sensitivity =  $TP/(FN+TP)$

<sup>f</sup> Specificity =  $TN/(FP+TN)$

<sup>g</sup> Accuracy =  $(TN+TP)/(TP+FP+TN+FN)$

<sup>\*</sup> KCED: Kurtosis Compensated Euclidean Distance.

second highest LR (+) value (12.6) and the lowest LR (-) value at 0.2. Although, 3-D SGM had the lowest sensitivity value at 97.7% and highest LR (+) value at 16.7%, it had the lowest sensitivity at just 38.5% and a low accuracy value of just 84.2%. On the whole, the KCED method out performed all other methods in terms of sensitivity , specificity and accuracy.

### 5.1.5 Likelihood ratios

The KCED method had a higher LR(+) versus the merged method. This indicated that the former method had a greater diagnostic weight in defending a positive test for CML (i.e., presence of the disease). In addition, the lower LR(-) value for the KCED method demonstrated its ability to defend a negative test for CML (i.e., rule out the presence of the disease) (Table 5.2). The likelihood ratios have shown that the KCED method is superior as compared to the other methods discussed in the thesis.

**Table 5.2** Likelihood ratios of the KCED method.

Method	LR(+) <sup>a</sup>	LR(-) <sup>b</sup>
KCED <sup>*</sup>	12.6	0.2

<sup>a</sup> LR(+): Positive Likelihood Ratio = Sensitivity/(1-Specificity)

<sup>b</sup> LR(-): Negative Likelihood Ratio = (1-Sensitivity)/Specificity

<sup>\*</sup> KCED: Kurtosis Compensated Euclidean Distance.

## 5.2 Discussion

In chapter 2, it was found that the Euclidean method suffered from low specificity, accuracy and LR(-) values [44]. This was because the algorithm computed the shortest distance in a multi-dimensional space. For non-symmetric distributions, the Euclidean distance failed to give a measure of the sample similarity from the PDF. Therefore to account for the non-symmetric nature of the PDF, KCED method was introduced in this chapter. KCED significantly increases the specificity, accuracy and LR(+) values, and reduces the LR(-) value, while maintaining sensitivity value at 81.8%. KCED was found to demonstrate the best performance values amongst all methods discussed so far, for the classification of CML versus healthy neutrophils from blood smear samples.



# Chapter 6

## Discussion

As technology combines spectroscopy with imaging, hyperspectral image processing has emerged as a niche which has found application in many fields like remote sensing, mineralogy, mapping vegetation cover and crop usage pattern etc. The exploitation of hyperspectral imaging started in the field of remote sensing where it was used to classify low spatial resolution hyperspectra; images captured from an aerial platform. This necessitated development of special algorithms to analyse the remotely sensed hyperspectral data sets [57-59]. However, use of high spatial and spectral resolution hyperspectral images of biological samples required different set of algorithms for classification. Unlike remotely sensed hyperspectral images, biomedical hyperspectral images are of high spatial and spectral resolutions. The resultant large data sets require special algorithms for processing and classification [44-45]. Many studies have ventured into the use of hyperspectral imaging on human blood cells. Greenspan *et al.* studied the use of spectral nuclear morphometry to group lymphocytes in lymphoproliferative disorders (i.e., chronic lymphatic leukemia, infectious mononucleosis and non-Hodgkin's lymphoma) [35]. However their work focused only on spectral data generated from the nuclear region of the cells. Another group studied spectral morphometric patterns in the nuclear region of chronic lymphocytic leukemia (B-CLL) cells and normal small lymphocytes [36]. Both these groups showed that when spectral data was combined with computa-



tional analyses, the patterns revealed a distinct correlation with the diseased versus the healthy cells. The research work focuses on extracting spectral and spatial information from the nuclear and cytoplasmic areas of the healthy and CML neutrophils.

From case reports, it is evident [1-3,6-8] that the detection of the presence of the Philadelphia chromosome alone is insufficient to mount symptoms in “preleukemic” patients, especially with those who have normal to mildly elevated WBCs. Hyperspectral images of cells capture the *in toto* information within each cell. Every hematological stain or dye has the potential to generate its own hyperspectral library of information, linked to the disease that is being diagnosed. In other words, hyperspectral imaging “fingerprints” the cells based on the data generated from the binding of the dye to the molecules within the healthy versus diseased cells. That is what we have shown with our studies on neutrophils.

In the present study various supervised methods for classification were explored and tested on the same set of samples, using uniform pre-processing methods. First two methods focused on classification using the spectral information content in hyperspectral image[44]. The next two methods were based on statistical distances [45]. They used the multidimensional Euclidean and the Mahalanobis distance to classify CML neutrophils. To improve the sensitivity of the Mahalanobis distance based classification method, it was merged with a novel method computed in the frequency domain. The last method focused on modifying the Euclidean distance method to take into account non-symmetric distributions. The detailed comparison of the all the methods based on their sensitivity, specificity and accuracy is given in Table. 6.1. The consolidated comparison of different methods based on their likelihood ratios is given in Table 6.2.

In chapter 2 the novel 3-D Spectral Gradient Mapping (3-D SGM) method was introduced [44]. The specificity (97.7%) and accuracy (84.2%) of classification was very high using this method. 3-D SGM exploits the inter-pixel and inter- band similarity between the training and the test pixels. However, the sensitivity values were found to be dramatically low. To improve the sensitivity measure of the clas-

**Table 6.1** Comparison of Sensitivity, Specificity and Accuracy Values.

Method	Classification parameters		
	Sensitivity <sup>a</sup>	Specificity <sup>b</sup>	Accuracy <sup>c</sup>
Windowed SAM <sup>*a</sup>	61.5	72.7	70.2
3-DSGM <sup>*b</sup>	38.5	97.7	84.2
Euclidean Distance	81.81	69.56	71.93
Mahalanobis Distance	54.54	89.13	82.45
Merged (Mahalanobis Distance and FDNPM <sup>*c</sup> )	63.6	93.5	87.7
KCED <sup>*d</sup>	81.8	93.5	91.2

<sup>a</sup> LR(+): Positive Likelihood Ratio = Sensitivity/(1-Specificity)

<sup>b</sup> LR(-): Negative Likelihood Ratio = (1-Sensitivity)/Specificity

<sup>\*a</sup> Windowed SAM: Windowed Spectral Angle Mapping.

<sup>\*b</sup> 3-D SGM: 3-Dimensional Spectral Gradient Mapping.

<sup>\*c</sup> FDNPM: Frequency Domain Normal PDF Matching.

<sup>\*d</sup> KCED: Kurtosis Compensated Euclidean Distance.

**Table 6.2** Comparison of Likelihood ratios.

Method	LR(+) <sup>a</sup>	LR(-) <sup>b</sup>
Windowed SAM <sup>*a</sup>	2.26	0.53
3-DSGM <sup>*b</sup>	16.70	0.63
Euclidean Distance	2.68	0.26
Mahalanobis Distance	5.02	0.51
Merged(Mahalanobis Distance and FDNPM <sup>*c</sup> )	9.80	0.40
KCED <sup>*d</sup>	12.60	0.20

<sup>a</sup> LR(+): Positive Likelihood Ratio = Sensitivity/(1-Specificity)

<sup>b</sup> LR(-): Negative Likelihood Ratio = (1-Sensitivity)/Specificity

<sup>\*a</sup> Windowed SAM: Windowed Spectral Angle Mapping.

<sup>\*b</sup> 3-D SGM: 3-Dimensional Spectral Gradient Mapping.

<sup>\*c</sup> FDNPM: Frequency Domain Normal PDF Matching.

<sup>\*d</sup> KCED: Kurtosis Compensated Euclidean Distance.

sification without significantly affecting the specificity and the accuracy values, the Euclidean and Mahalanobis distances were used to identify and group CML from healthy neutrophils [44] in chapter 3. The Mahalanobis distance method had higher specificity (89%) and accuracy (82.5%) performance values as compared to the results from the Euclidean distance algorithm which had specificity value of 69.6% and accuracy value of 71.9%, respectively. As far as sensitivity was concerned, the Euclidean distance method had a 18.2% chance of missing a positive diagnosis versus 45.5% for the Mahalanobis distance approach. On the whole, neither method was stronger for all three performance parameters.

To improve the results, in chapter 4, a combined approach using Mahalanobis distance method merged with a novel classification method, FDNPM was explored. FDNPM compares the test pixel to the PDF of the training pixel set in the frequency domain. Both the Mahalanobis distance and the FDNPM methods used PDF for deriving the classification metrics. The Mahalanobis distance method computes the similarity index by calculating the Mahalanobis distance of a test pixel from the PDF of the training pixel to classify it. However, the whole neutrophil cell *per se* is represented by a set of pixels. To classify them correctly, it is necessary to compute its similarity with the set of training pixels. The FDNPM method computes the similarity by considering the entire set of test pixels of the nucleus and cytoplasm of the neutrophils. Further, the Fourier transform of the hyperspectral cube represents the information content as sum of scaled frequency responses. The frequencies gave the rate of change of information over the space, i.e., along  $X$  (the rows),  $Y$  (the columns) and  $Z$  (the wavelength bands) axes. Therefore, by matching the PDF in the frequency domain, the spatial rate of information content change of the hyperspectral cube was compared. The Mahalanobis distance was merged with the FDNPM method to improve on the sensitivity value without compromising the former's high specificity and accuracy values. In fact, the merged method was found to have increased the specificity (by 4.4%) and accuracy (by 5.2%) values as well. The merged method has much higher LR(+) values (9.80) than solely using the

Mahalanobis distance approach (5.02). With a lower LR(-) value (0.40), FDNPM method improvises the Mahalanobis distance method by reducing the likelihood of missing diseased cases compared to only using the Mahalanobis distance; LR(-) = 0.51; [45,48,62].

As discussed in chapter 3, the Euclidean method suffered from low specificity, accuracy and LR(-) values [45]. This was because the algorithm computed the shortest distance in a multi-dimensional space. For non-symmetric distributions, the Euclidean distance fails to give a measure of the sample similarity from the PDF. Therefore to account for the non-symmetric nature of the PDF, KCED method was introduced in chapter 5. KCED significantly increases the specificity, accuracy and LR (+) values and reduces the LR(-) value, while maintaining sensitivity value at 81.8%. Therefore, KCED was found to have produced the best results when compared with all other methods discussed in this thesis, for classification of CML versus healthy neutrophils from blood smear samples [45,62].

The research work focused on whether we could use hyperspectral imaging to distinguish intra-cell differences of healthy versus diseased states. The inclusion criteria for healthy blood donors in the present research work were very stringent [44,45,62]. Wang *et al.* used machine learning approaches successfully to differentiate WBCs (i.e., inter-cell comparison). But they had to use 70% of their samples for training [50]. The present thesis is about research which focuses on intra-cell differences and uses less than 4% of the samples for training.

The six methods described in the thesis for classifying CML neutrophils from the hyperspectral images of their blood samples have been adequately analysed in terms of their ability to correctly classify the blood sample as diseased or non-diseased. However, the actual use of the methods depends on how efficiently they can be implemented. This needs the algorithms which form part of these classification methods to be computationally viable in terms of use of the processor and time taken to execute. This can be gauged based on the computational complexity of the classification methods/ algorithms. The six methods described in the thesis

have been evaluated and listed below in increasing order of their computational complexity.

- Windowed Spectral Angle Mapping (WSAM) method
- Euclidean distance method
- Mahalanobis distance method
- Kurtosis Compensated Euclidean Distance (KCED) method
- 3- Dimensional Spectral Gradient Mapping (3-D SGM) method
- Merged Method comprising of Mahalanobis distance method and Frequency Domain Normal Probability Matching (FDNPM) method

Overall collective analysis of the six methods based on their effectiveness for blood cell classification and efficiency in terms of computational complexity, tells us that the Kurtosis Compensated Euclidean Distance (KCED) method is the best method which provides highest sensitivity, accuracy and lowest negative likelihood ratio. Also, unlike other methods, KCED achieves a remarkable trade off between sensitivity and specificity, as well as between positive and negative likelihood ratios, by providing the second highest specificity and positive likelihood ratio values. In terms of computational complexity, it is better than 3-D SGM and Merged methods, while being marginally more complex than the Mahalanobis distance method for classification of CML neutrophils from hyperspectral images.



# Chapter 7

## Conclusion

The routine diagnosis for leukemia is based on the visual assessment of a stained blood smear sample using a light microscope. Leukemia can be broadly classified as Acute Lymphoblastic Leukemia (ALL) and Chronic myeloid leukemia (CML). In this thesis, we have explored the various methods for classifying CML based on the hyperspectral image cubes of blood samples. CML is a type of cancer that originates in the bone marrow, where blood cells are created. It is one of the most common blood cancers and takes a longer time to develop. It affects adults more than children and the risk of developing CML increases after the age of 65. The presence of leukemia cells is characterized by an unusual increase in the number of white blood cells [9]. Neutrophils are the most abundant (60 - 70%) cell type in blood circulation and perform an important role in the ensuring innate immunity in human body by providing the first line of defense [16-17]. Therefore, the study has appropriately focused on the aspect of detecting CML based on the affected neutrophils.

Advances in medical imaging technologies have revolutionized health care delivery globally by making diagnostics more efficient through automation. These advances have helped the diagnosis go beyond limited detection capabilities of the human eye. One such technique is the use of hyperspectral images. As discussed above, hyperspectral images hold immense potential to revolutionize the way health

care is delivered world wide. The huge amount of information content in hyperspectral image cubes needs to be effectively exploited in order to achieve accurate diagnosis. The present research work utilizes this capability of the hyperspectral images to ensure accurate diagnosis of CML affected neutrophils from the blood samples.

Hyperspectral image processing is a widely used imaging technique in the field of remote sensing, where it is used to classify objects from low spatial resolution but high spectral resolution images. But with advent of better imaging technology, it has become possible to capture high spatial as well as high spectral resolution images. This has opened the scope of using hyperspectral images for diagnostic applications in the field of medical imaging. However, due to the high resolution, these image cubes generate large datasets and require efficient algorithms to extract useful information out of these huge datasets. The existing hyperspectral image processing techniques, which was applied on remotely sensed low-resolution datasets, cannot be applied on high resolution biomedical images like blood smears. Therefore, the research work described in the thesis has explored six innovative methods which can be employed to use hyperspectral data sets to diagnose CML from blood samples.

## **7.1 Other Medical Applications of Hyperspectral Image Processing**

Hyperspectral images have immense potential in the field of medical applications. Specifically, being a high resolution imaging technique, its application is mostly in the field of medical diagnostics. This includes analysis of hyperspectral image cubes of various types of biomedical tissues and samples. The present thesis explores innovative methods for analysing blood samples to classify CML neutrophils.



### **7.1.1 Hyperspectral Imaging for Analysing Biomedical Tissues**

Hyperspectral imaging can be applied to any biological tissue to examine it at the molecular level and extract unique information about the state of the tissue based on the tissue and wavelength interaction. Some potential bio-medical applications of hyperspectral imaging are listed below.

- Diagnosis of melanoma from hyperspectral images of the skin.
- Alzheimer's Disease by detecting Amyloid beta from the hyperspectral image of the retina.
- In-vivo colon detection based on hyperspectral endoscopy of the colon.
- Determination of brain tumour margins based on hyperspectral image of the tumour.
- Diagnosis of blood abnormalities based on hyperspectral image of blood samples.

### **7.1.2 Hyperspectral Imaging of Blood Samples**

If the standard operating protocol is maintained across different blood-based cancers, i.e., blood smear preparations and staining protocol, there should be sufficient differences that can be picked up using hyperspectral imaging to clearly differentiate the different cancers. This sensitive technology picks up signals at the molecular level using the visible as well as near infra red range of the spectrum. Each different blood-based cancer has its own molecular signature which is currently used at the clinical level. These blood cancers may share some common hyperspectral image signals, even with healthy blood cells. The main objective is to look for those

wavelengths that differentiate them, just like the differences seen between CML and healthy neutrophils. As for treatment efficacy, some clinical data pertaining to the patient can be merged into the analysis. It will not be only the treatment per se. Once a retrospective library or data is compiled, there is potential to be able to predict treatment efficacy in prospective patients once the patient data is added.

## **7.2 Challenges in Use of Hyperspectral Images for Medical Diagnosis**

While the use of hyperspectral images to diagnose diseased leukemia cells like neutrophils is an innovative and path breaking approach towards low cost diagnostics, it poses some inherent challenges as well. The first challenge encountered in use of hyperspectral imaging for medical diagnosis is the inability of the hyperspectral imager to capture 3-D image of the sample. This means that, while the surface of the tissue can be captured and analysed with great detail in various spectral bands, the inner cells of the tissue cannot be captured without cutting through the surface. Therefore, hyperspectral imaging for non invasive diagnosis of 3-D biomedical samples is limited. However, in this thesis the hyperspectral images were captured in the visible and near infra red bands. The penetration power of the the wavelengths in the near infra red bands is more than the visible band. Therefore, the information content from the tissue present under the surface could be obtained from the wavelengths in the near infra red bands. Thus the first challenge could be partially offset by capturing the hyperspectral image in the near infra red band. The second challenge is the high cost of the hyperspectral imaging equipment. Hyperspectral sensors which are used in the imager and the associated optics are expensive. Secondly, being a niche technology, especially in the field of medical health care, the availability of trained manpower to undertake precise operation of the imaging system is limited. Also the procedure for capturing hyperspectral images requires a

fair knowledge of spectroscopy. Finally, hyperspectral image cubes are humongous datasets with very high information content. Effective and efficient algorithms need to be developed specific to the intended application in order to exploit the hyperspectral datasets. The present study as described in the thesis, has proposed six such innovative methods for classifying CML affected neutrophils from hyperspectral images of their blood samples and compared them based on their ability to correctly classify the blood cells and also in terms of their computational complexity.

### 7.3 Future Scope of Work

The existing work concentrates on CML diagnosis by identifying the affected neutrophils using hyperspectral images. Use of hyperspectral images to detect other affected blood cells like leukocytes has the potential to improve the diagnostic efficacy of the methods described.

For computational efficiency, PCA has been used for dimension reduction to exclude the redundant wavelengths. Other methods of dimension reduction like minimum noise fraction analysis, independent component analysis etc, may also be explored.

Hyperspectral imaging offers high spectral resolution. However, the spatial information is limited to two dimensions by virtue of the limitation of the hyperspectral sensors used in clinics to capture 3-D information. Significant advances in the field of optical systems and spectroscopy are being made to devise an implementable technology that can provide 3-D quantitative depth information of biomedical tissues. This would greatly enhance the diagnostic capabilities of the hyperspectral imaging systems in medical applications. Till this technology matures completely, hyperspectral imaging can be used in combination with classical 3-D imaging techniques like CT and MRI scans for better analysis of bio-medical tissue samples in order to improve diagnostic capabilities.

The present study proposes novel supervised algorithms based on available sam-

ples to classify the blood samples from their hyperspectral images. Advanced machine learning and deep learning based unsupervised algorithms can also be implemented over hyperspectral image cubes to automate classification. The more the number of layers in the learning algorithms, the better is the classification. However, it would also increase the compute requirement due to increase in computational complexity of the method for achieving similar results. Therefore, the use of advanced machine learning and deep learning algorithms may be explored for classification based on the application desired.



# Bibliography

- [1] P. C. Nowell and D. A. Hungerford, “A minute chromosome in human chronic granulocytic leukemia,” *Science*, vol. 142, pp. 1497, 1960.
- [2] P. C. Nowell, “Discovery of the Philadelphia chromosome: a personal perspective,” *J. Clin. Invest.*, vol. 117, no. 8, pp. 2033-2035, Aug. 2007, DOI: 10.1172/JCI31771.
- [3] J. D. Rowley, “A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescene and Giemsa staining,” *Nature*, vol. 243, no. 5405, pp. 290-293, Jun. 1973, DOI: 10.1038/243290a0.
- [4] B. Chereda, J.v. Melo, “NAtural course and biology of CML,” *Ann Hematol.*, Vol. 94 Suppl 2:S107-21, Apr. 2015, DOI: 10.1007/s00277-015-2325-z.
- [5] S. Faderl, M. Talpaz, Z. Estrov, S. O’Brien, R. Kurzrock, and H. M. Kantarjian, “The biology of chronic myeloid leukemia,” *N. Engl. J. Med.*, vol. 341, no. 3, pp. 164-172, Jul. 1999, DOI: 10.1056/NEJM199907153410306.
- [6] S. D. Hudnall, J. Northup, N. Panova, K. Suleman, and G. Velagaleti, “Prolonged leukemic phase of chronic myelogenous leukemia,” *Exp. Mol. Pathol.*, vol. 83, no. 3, pp. 484-489, Dec. 2007, DOI: 10.1016/j.yexmp.2007.08.006.
- [7] L. L. Aye *et al.*, “Preleukemic phase of chronic myelogenous leukemia: morphologic and immunohistochemical characterization of 7 cases,” *Ann. Diagn. Pathol.*, vol. 21, pp. 53-58, Apr. 2016, DOI: 10.1016/j.anndiagpath.2015.12.004.

- [8] J. W. Kuan, A. T. Su, C. F. Leong, M. Osato, and G. Sashida, "Systematic review of pre-clinical chronic myeloid leukaemia," *Int. J. Hematol.*, vol. 108, no. 5, pp. 465-484, Nov. 2018, DOI: 10.1007/s12185-018-2528-x.
- [9] B. J. Druker "Imatinib as a paradigm of targeted therapies," *Adv Cancer Res.* vol. 91, pp. 1-30. DOI: 10.1016/S0065-230X(04)91001-9. PMID: 15327887.
- [10] H. Menon, " Non-adherence to CML therapy and its clinical implications in India," *Natl Med J India.*, Vol 30(3), May-Jun 2017, pp. 142-147.
- [11] S. Arunachalam, "Applications of machine learning and image processing techniques in the detection of leukemia," *Int. J. Sci. Res. Comput. Sci. Eng.* Vol. 8(2), 2020, pp. 77–82, DOI: <http://dx.doi.org/10.26438/ijsrcse/v8i2.7782>.
- [12] X. Zhang, K. Zhang, M. Jiang, and L. Yang, "Research on the classification of lymphoma pathological images based on deep residual neural network," *Technol. Health Care*, vol. 29, no. Suppl. 1, pp. 335-344, 2021, DOI: 10.3233/THC-218031.
- [13] S. Shafique, and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technol. Cancer Res. Treat.*, vol. 17, Jan. 2018, Art. no. 1533033818802789.
- [14] S. Mohapatra, D. Patra, and S. Satpathy, "An ensemble classifier system for early diagnosis of acute lymphoplastic leukemia in blood microscopic images," *Neural. Comput. & Applic.*, vol. 24, pp. 1887-1904, Jun. 2014, DOI: 10.1007/s00521-013-1438-3.
- [15] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, Third Quarter 2006, DOI: 10.1109/MCAS.2006.1688199.
- [16] J.T. Dancey, K.A. Deubelbeiss, L.A. Harker, C.A. Finch, "Neutrophil kinetics in man," *J. Clin. Investig.*, vol. 58, 1976, pp. 705–715, doi:10.1172/JCI108517

- [17] C. Rosales, “Neutrophil: a cell with many roles in inflammation or several cell types?” *Front. Physiol.*, vol. 9, 113, Feb. 2018, DOI: 10.3389/fphys.2018.00113.
- [18] B.K. Mohan, A. Porwal, ”Hyperspectral image processing and analysis,” *Curr. Sci.*, Vol. 108(5), 2015, pp. 833–841.
- [19] P. K. Varshney, and M. K. Arora, *Advanced image processing techniques for remotely sensed hyperspectral data*. Berlin Heidelberg, Germany: Springer, 2004. DOI: 10.1007/978-3-662-05605-9.
- [20] R. Levenson, J. Beechem, G. McNamara, ”Spectral imaging in preclinical research and clinical pathology,” *Anal. Cell. Pathol.*, 35 (5–6), 2012, pp. 339–361, DOI: 10.3233/ ACP-2012-0062
- [21] A. Panda, and D. Pradhan, “Hyperspectral image processing for target detection using spectral angle mapping.” *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 1098–1103, DOI: 10.1109/IIC.2015.7150911.
- [22] M. A. Cochrane, and C. M. Souza Jr., “Linear mixture model classification of burned forests in the Eastern Amazon,” *Int. J. Remote Sensing*, vol. 19, no. 17, pp. 3433–3440, Nov. 1998, DOI: 10.1080/014311698214109.
- [23] H. Ren, and C-I. Chang, “Automatic spectral target recognition in hyperspectral imagery,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003, DOI: 10.1109/TAES.2003.1261124.
- [24] J. A. Richards, and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin Heidelberg, Germany: Springer-Verlag, 2006. DOI: 10.1007/3-540-29711-1.
- [25] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*. New Jersey, USA: John Wiley & Sons, 2015.



- [26] J. R. Carr, and K. Matanawi, “Correspondence analysis for principal components transformation of multispectral and hyperspectral digital images,” *Photogramm. Eng. Remote Sensing*, vol. 65, no. 8, pp. 909–914, Aug. 1999.
- [27] Q. Du, and J. E. Fowler, “Low-complexity principal component analysis for hyperspectral image compression,” *Int. J. High Perform. Comput. Appl.*, vol. 22, no. 4, pp. 438–448, Nov. 2008, DOI: 10.1177/1094342007088380.
- [28] C. Rodarmel, and J. Shan, “Principal component analysis for hyperspectral image classification,” *Surv. Land Inf. Syst.*, vol. 62, no. 2, pp. 115–123, Jun. 2002.
- [29] E. C. Meidunas, “Robust estimation of Mahalanobis distance in hyperspectral images,” Ph.D. dissertation, Sch. Eng. Manag., Air Force Institute of Technology, Air Univ., Wright-Patterson AFB, Ohio, USA, 2006.
- [30] M. Imani, “Difference-based target detection using Mahalanobis distance and spectral angle,” *Int. J. Remote Sensing*, vol. 40, no. 3, pp. 811–831, 2019, DOI: 10.1080/01431161.2018.1519280.
- [31] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation.*, New Jersey, USA: John Wiley & Sons, 2015.
- [32] Y. Garini, I. T. Young, and G. McNamara, “Spectral imaging: principles and applications,” *Cytometry Pt A*, vol. 69A, no. 8, pp. 735–747, Sep. 2006, DOI: 10.1002/cyto.a.20311.
- [33] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, “Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 5 (2), 2012, pp. 354–379, DOI: 2194696.
- [34] A. Majda, R. Wietecha-Posłuszny, A. Mendys, A. Wójtowicz, and B. Łydzba-Kopczyńska, “Hyperspectral imaging and multivariate analysis in the dried

- blood spots investigations,” *Appl. Phys. A*, vol. 124, Mar 2018, Art. no. 312, DOI: 10.1007/s00339-018-1739-6.
- [35] H. Greenspan, C. Rothmann, T. Cycowitz, Y. Nissan, A. M. Cohen, and Z. Malik, “Classification of lymphoproliferative disorders by spectral imaging of the nucleus,” *Histol. Histopathol.*, vol. 17, no. 3, pp. 767-773, Jul. 2002, DOI: 10.14670/HH-17.767.
- [36] Z. Malik, C. Rothmann, T. Cycowitz, Z. J. Cycowitz, and A. M. Cohen, “Spectral morphometric characterization of B-CLL cells versus normal small lymphocytes,” *J. Histochem. Cytochem.*, vol. 46, no. 10, pp. 1113-1118, Oct. 1998, DOI: 10.1177/002215549804601003.
- [37] K. Masood, and N. Rajpoot, “Spatial analysis for colon biopsy classification from hyperspectral imagery,” *Annals of the BMVA*, vol. 2008, no. 4, pp. 1-16, Jan. 2008.
- [38] Q. Wang, J. Wang, M. Zhou, Q. Li, and Y. Wang, “Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperpsectral imaging technology,” *Biomed. Opt. Express*, vol. 8, no. 6, pp. 3017–3028, May. 2017, DOI: 10.1364/BOE.8.003017.
- [39] G. Lu, D. Wang, X. Qin, L. Halig, S. Muller, H. Zhang, A. Chen, B.W. Pogue, Z.G. Chen, B. Fei, ”Framework for hyperspectral image processing and quantification for cancer detection during animal tumor surgery,” *J. Biomed. Opt.*, 20 (12), 2015, DOI: 126012.
- [40] P.K. Varshney, M.K. Arora, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data.*, Berlin Heidelberg: Springer, 2004, DOI: <http://dx.doi.org/10.1007/978-3-662-05605-9>.
- [41] G. Lu, B. Fei, ”Medical hyperspectral imaging: a review,” *J. Biomed Opt.*, 19(1), 2014, DOI: 010901.

- [42] D. Manolakis, D. Marden, G.A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Linc. Lab. J.*, 14 (1), 2003, pp. 79–116.
- [43] C. Dai, Q. Li, J. Liu, "Blood cells classification using hyperspectral imaging technique," *J. Bioinf. Biol. Eng.*, 1 (1), 2013, pp. 27–33.
- [44] A. Panda, R. B. Pachori, N. Kakkar, M. J. John, and N. D. Sinnappah-Kang, "Screening chronic myeloid leukemia neutrophils using a novel 3-Dimensional Gradient Mapping algorithm on hyperspectral images," *Comput. Methods Programs Biomed.*, vol. 220, Jun. 2022, Art. no. 106836, DOI: 10.1016/j.cmpb.2022.106836.
- [45] A. Panda, R. B. Pachori, and N. D. Sinnappah-Kang, "Classification of chronic myeloid leukemia neutrophils by hyperspectral imaging using Euclidean and Mahalanobis distances," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 103025, DOI: 10.1016/j.bspc.2021.103025.
- [46] A. Karami, M. Yazdi, G. Mercier, "Compression of hyperspectral images using discrete wavelet transform and tucker decomposition," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 5 (2), (2012), pp. 444–450.
- [47] R. Trevethan, "Sensitivity, specificity, and predictive values: Foundations, plia-bilities, and pitfalls in research and practice," *Front. Public Health*, vol. 5, Nov. 2017, 307, DOI: 10.3389/fpubh.2017.00307.
- [48] F. Habibzadeh, and P. Habibzadeh, "The likelihood ratio and its graphical representation," *Biochem. Med. (Zagreb)*, vol. 29, no. 2, Jun. 2019, Art. no. 020101, DOI: 10.11613/BM.2019.020101.
- [49] S. McGee, "Simplifying likelihood ratios," *J. Gen. Intern. Med.*, vol. 17, no. 8, pp. 647–650, Aug. 2002, DOI: 10.1046/j.1525-1497.2002.10750.x.
- [50] Q. Wang, J. Wang, M. Zhou, Q. Li, Y. Wen, and J. Chu, "A 3D attention networks for classification of white blood cells from microscopy hyperspec-

- tral images.” *Opt. Laser Technol.*, vol. 139, Jul. 2021, Art. no. 106931, DOI: 10.1016/j.optlastec.2021.106931.
- [51] S. Mallet, S. Halligan, M. Thompson, G. S. Collins, and D. G. Altman, “Interpreting diagnostic accuracy studies for patient care,” *BMJ*, vol. 345, 2012, Art. no. e3999, DOI: 10.1136/bmj.e3999.
- [52] H. Mark, J. Workman Jr, *Chemometrics in Spectroscopy.*, Cambridge: Academic Press, 2018, DOI: <http://dx.doi.org/10.1016/C2015-0-04023-0.F>.
- [53] G. Girouard, A. Bannari, A. El Harti, A. Desrochers, ”Validated spectral angle mapper algorithm for geological mapping: comparative study between Quickbird and Landsat-TM,” in: O. Altan (Ed.), *Proceedings of the XXth ISPRS Congress.*, Istanbul, Turkey, 2004, pp. 599–604.
- [54] A. J. Steven, J. Zhuo, and E. R. Melhem, “Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain,” *AJR Am. J. Roentgenol.*, vol. 202, no. 1, pp. W26-33, Jan. 2014, DOI: 10.2214/AJR.13.11365.
- [55] A. Arab, A. Wojna-Pelczar, A. Khairnar, N. Szabó, and J. Ruda-Kucerova, “Principles of diffusion kurtosis imaging and its role in early diagnosis of neurodegenerative disorders,” *Brain Res. Bull.*, vol. 139, pp. 91-98, May. 2018, DOI: 10.1016/j.brainresbull.2018.01.015.
- [56] R. N. Henriques *et al.*, “Diffusional kurtosis imaging in the diffusion imaging in Python Project,” *Front. Hum. Neurosci.*, vol. 15, Jul. 2021, Art. no. 675433, DOI: 10.3389/fnhum.2021.675433.
- [57] S.S. Chiang, C.I. Chang, I.W. Ginsberg, ”Unsupervised target detection in hyperspectral images using projection pursuit,” *IEEE Trans. Geosci. Remote Sens.*, 39 (7), 2001, pp. 1380–1391, DOI: 934071.

- [58] S. Yang, Z. Shi, W. Tang, "Robust hyperspectral image target detection using an inequality constraint," *IEEE Trans. Geosci. Remote Sens.*, 53 (6), 2014, pp. 3389–3404, DOI: 2375351.
- [59] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H. Goetz, "The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, 44 (2–3), 1993, pp. 145–163, DOI:10.1016/0034-4257(93)90013-N.
- [60] L.Wang, Y. Zhang, J. Feng, "On the euclidean distance of images, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1334-1339, 2005, DOI: <https://doi.org/10.1109/TPAMI.2005.165>.
- [61] N. Srivastava, and S. Rao, "Learning-based text classifiers using the Mahalanobis distance for correlated datasets," *Int. J. Big Data Intell.*, Vol. 3, no. 1, 2016, pp. 18-27, DOI: 10.1504/IJBDI.2016.073901.
- [62] A. Panda, R.B. Pachori, N. Kakkar, M.J John, N.D. Sinnappah-Kang,"The Kurtosis compensated Euclidean distance method for screening and early detection of chronic myeloid leukemia neutrophils from hyperspectral images," ., Vol

# List of publications

## Publications from thesis

Journal papers:

1. A. Panda, R. B. Pachori, and N. D. Sinnappah-Kang, "Classification of chronic myeloid leukemia neutrophils by hyperspectral imaging using Euclidean and Mahalanobis distances," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 103025, DOI: 10.1016/j.bspc.2021.103025.
2. A. Panda, R. B. Pachori, N. Kakkar, M. J. John, and N. D. Sinnappah-Kang, "Screening chronic myeloid leukemia neutrophils using a novel 3-Dimensional Gradient Mapping algorithm on hyperspectral images," *Comput. Methods Programs Biomed.*, vol. 220, Jun. 2022, Art. no. 106836, DOI: 106836.