Classification of White Blood Cells from Microscopic Images

Author

Muhammad Abbas Hussain

Registration Number

171588

Supervisor

Dr. Arslan Shoukat

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

February, 2019

Classification of Sub-White Cells from Microscopic Images

Author

Muhammad Abbas Hussain

Registration Number

171588

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor:

Dr. Arslan Shoukat

Thesis Supervisor's Signature: _____

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

February, 2019

Declaration

I certify that this research work titled "*Classification of Sub-White Cells from Microscopic Images"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Muhammad Abbas Hussain

2016-NUST-Ms-CE-00000171588

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

Muhammad Abbas Hussain

Registration Number

00000171588

Signature of Supervisor

Acknowledgements

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought which You setup in my mind to improve it. Indeed I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life.

I would also like to express special thanks to my supervisor Dr. Arslan Shoukat for his help throughout my thesis and also for Patter Recognition & Analysis which he has taught me during master coursework. I can safely say that I haven't learned any other engineering subject in such depth than the ones which he has taught.

I would also like to pay special thanks to Ibtihaj Ahmed Khan for his tremendous support and cooperation. Each time I got stuck in something, he came up with the solution. Without his help I wouldn’t have been able to complete my thesis. I appreciate his patience and guidance throughout the whole thesis.

I would also like to thank Dr. Usman Akram and Dr. Farhan Hussain for being on my thesis guidance and evaluation committee.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional parents whose tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Image processing has become an important pillar in modern age of technology. It plays a significant role in medical field. It makes the diseases diagnosis and analysis more accurate and speedup this process. Human health is usually measured from the immune system of body which is natural defense system against infections and invaders. Human immune system contain white blood cells (WBC's) which are good indicator of many diseases like bacterial infections, AIDS, cancer, spleen, etc. White blood cells have been further classified into four major classes such that monocytes, lymphocytes, eosinophils, neutrophils on the base of their nucleus, shape and cytoplasm. Traditionally during tests in laboratories, pathologist and hematologist analysis of these cells in blood is done through microscope and then classified manually as normal or abnormal. Counting is also done manually. This hard work takes more time and increase the chance of human error. In last decade, some research has been done on this field of medical science in order to automate the process. In this research, a method is introduced to automatically classify the sub types of white blood cells and compare the results to previous techniques. The proposed method includes: preprocessing, color pallet based segmentation, hybrid features (pattern and shape based features) extraction and neural network. The classification accuracy of 96.5% is achieved through the proposed method.

Key Words: *White Blood Cells (WBC's), AIDS, Monocytes, Lymphocytes, Eosinophils, Neutrophils, Segmentation, Neural Network (NN)*

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym Used	Definition
AIDS	Acquired Immune Deficiency Syndrome
CNN	Convolution Neural Network
FPR	False Positive Rate
GBPS	Genetic-Based Parameter Selector
GLCM	Grey Level Co-Occurrence Matrix
HSV	Hue, Saturation, Value
JPEG	Joint Photographic Experts Group
LBP	Local Binary Pattern
LNE	Leukocyte Nucleus Enhancer
NN	Neural Network
PCA	Principal Component Analysis
RGB	Red, Green, Blue
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SVM	Support Vector Machine
TPR	True Positive Rate

Chapter 1 Introduction

White blood cell's classification and their count has significant role in the process of diagnosis of a disease and analysis of human health. Recently image processing has great worth in medical field to make ease and automate the medical diagnosis process. This chapter give a brief overview of some of the preliminary background, which include basic human immune system, introduction of pathology, background, scope and motivation.

1.1 Human Immune System

Human body immune system is natural defense system against infections and invading organism. It is made up of antibodies, white blood cells, tissue and proteins, which fight against those bacteria's and viruses that are harmful for human health. Normally, immune system plays important role to maintain the body health. Any disorders of human immune system lead to infection or illness.

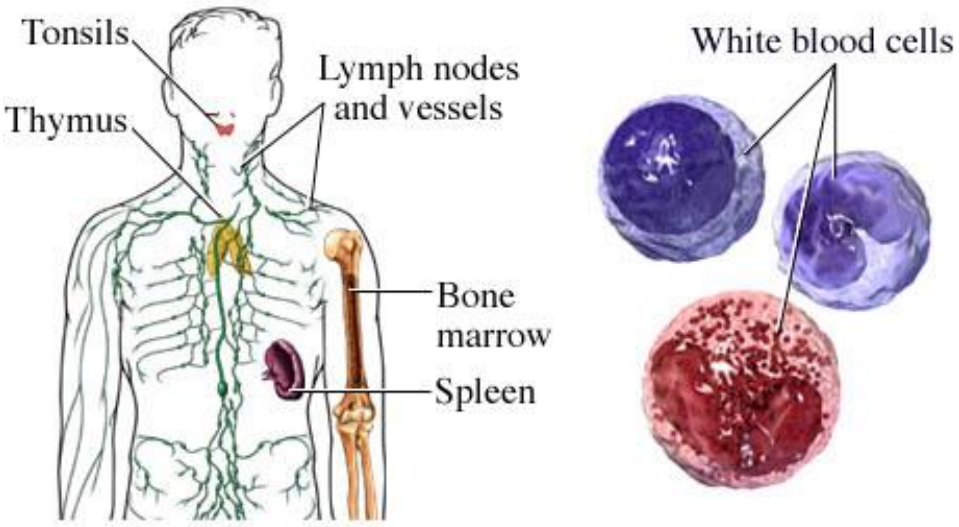


Figure 1.1 Human Immune System [1]

Immune system is identical to a network of cells and organs which protect the body. White blood cells play important role in immune system which known as leukocytes. Leukocytes resides at many parts of the body such that spleen, thymus and bone marrow as shown in Figure 1.1. Due to high percentage, it is also known as lymphoid organ. A clumps of lymphoid tissues

are available throughout the body in which leukocytes resides. It circulates all over the body through lymphatic vessels and blood vessels. Leukocytes mostly produce in the red marrow of bones and little quantity at other special glands. Normally, a health human body contain 4 to 11 thousands cells in one cubic inch blood. Five major types of white blood cells are given below and shown in Figure 1.2.

- Neutrophil
- Lymphocyte
- Monocyte
- Eosinophil
- Basophil

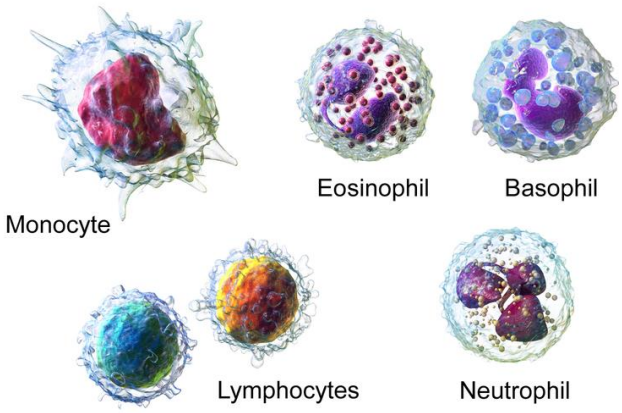


Figure 1.2 White Blood Cells [2]

1.1.1 Neutrophil

Neutrophil is sub type of white blood cells, which is present in a majority percentage. A healthy person has 50 to 70 percent neutrophil of the white blood cells. It has multi-lobed nucleus which has normally two to five lobes are shown in Figure 1.3. Neutrophil cell's cytoplasm has light pink in color and it has also granules which are in numerous and in small size. It is front line defender when infections strike. It defends against fungal and bacterial infection. Its granules become darker when it fights against the infection. The life spans of neutrophil cell nearly 5.4 days.

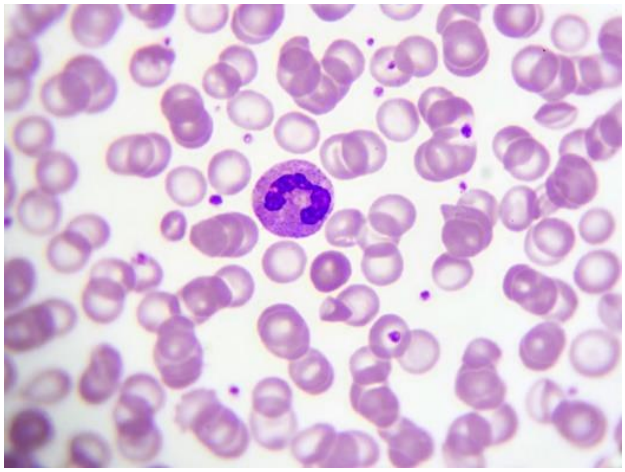


Figure 1.3 Neutrophil

1.1.2 Lymphocytes

Lymphocytes are present in second highest percentage in white blood cells, which are 20 to 40 percent in an adult human. It has mono-lobed nucleus as shown in Figure 1.4. Its shape can vary but under normal condition its size is small and has smooth round shape which contain a small amount of blueish cytoplasm. A special characteristic of this cell is that it has a memory, it has ability to recognize the foreign invader which attacked in past. Next time when attack the same invader, it works against fast as compare to previous. When it fights against the invader, its size will be large and more cytoplasm in it. Normally, it defends against viruses and bacteria's.

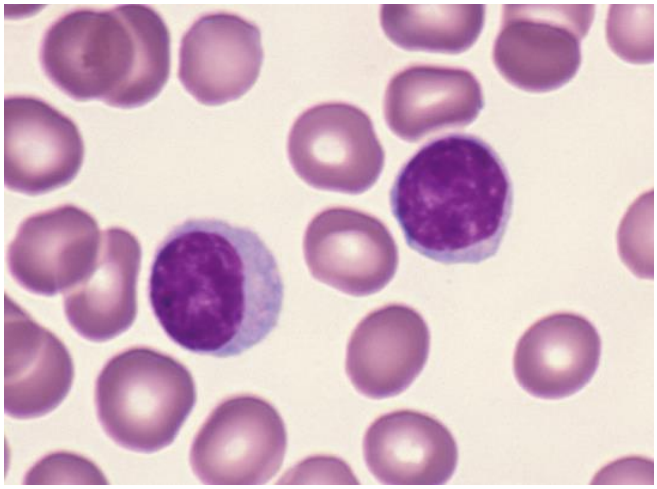


Figure 1.4 Lymphocytes

1.1.3 Monocytes

Monocytes are not present in high percentage in blood. Its percentage in adult human body is up to 5.3 percent. It seems to look like lymphocytes but it has large in size and more cytoplasm. Its shape not smoothly and the color of cytoplasm is dull blueish gray as shown in figure 1.5. Granules also in it which are lightly stain. One thing more, vacuoles are also in it which looks like to be holes in cytoplasm. Monocytes not attacked against the invaders like to other white blood cells, but it provides defense by digesting the foreign particles. Life span of monocytes is between hours to days.

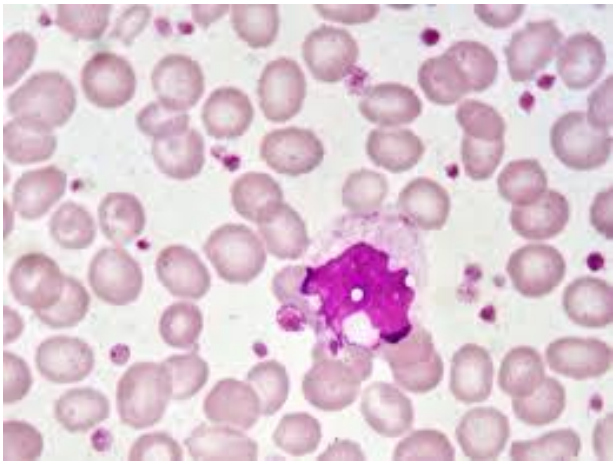


Figure 1.5 Monocytes

1.1.4 Eosinophils:

Eosinophil percentage in human is 2 to 4 percent which varies in a day or seasonally. It has bilobed nucleus which is attached to a thin strand. Its cytoplasm is full of granules and those granules have pink orange color which shown in Figure 1.6. Eosinophil responded against parasitic infections, allergies, disease of the spleen, central nervous system and collagen diseases. It secretes a type of chemical which fight against the parasites such that hooks worms and tape worms. It also provides defense from allergies such as hay fever, asthma and hives. Life span of eosinophils is 8 to 12 days.

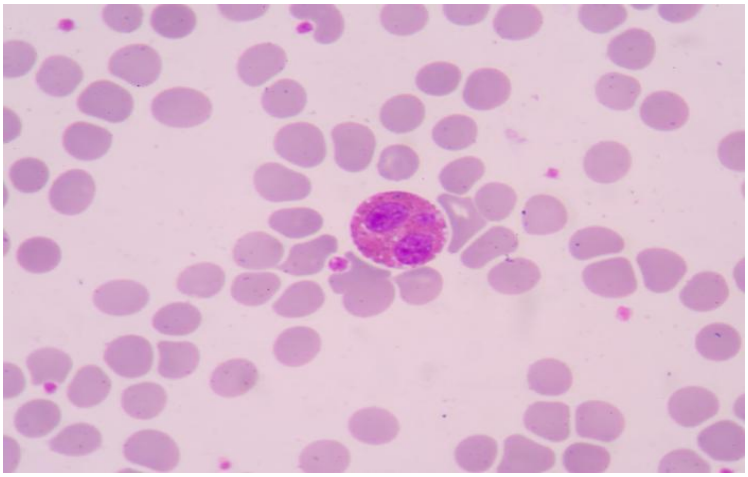


Figure 1.6 Eosinophils

1.1.5 Basophil

Basophil is the least percentage of white blood cells in human body which contain less than 0.5 percent. It has bi or tri lobed nucleus and it's hard to see due to coarse granules in the cytoplasm. Its granules have deep blue purple color as shown in Figure 1.7. These smallest cells create alarm to immune system when invaders to attack the blood. It secretes a chemical such as histamine which indicates allergic disease. Its life span is only few hours.

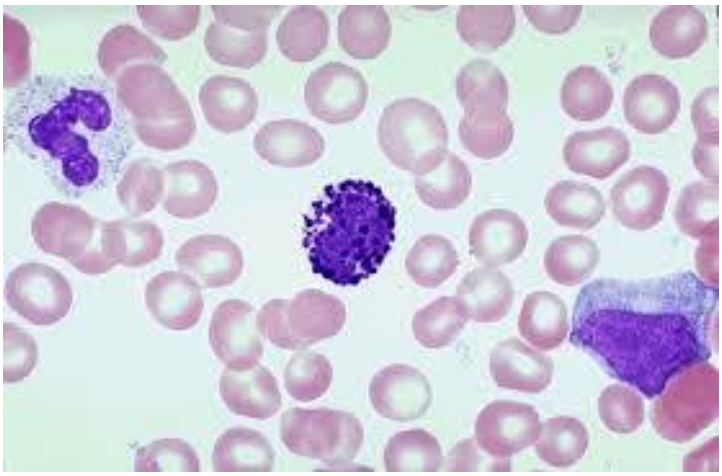


Figure 1.7 Basophil

Table 1-1 Summary Details of WBC's

Sr#	Type	Percentage (%)	Diameter (μm)	Nucleolus	Granules	Life Time
1	Neutrophil	50 - 70 %	10 -12	Multilobed	Fine, Faintly Pink	5.4 days
2	Lymphocyte	20 - 40 %	7- 8	Single	None	Number of years
3	Monocyte	5.3 %	15 - 30	Single	None	Hours to Days
4	Eosinophil	2.3 %	10 – 12	Bi-lobed	Full of pink-orange	8 – 12 Days
5	Basophil	0.4 %	12 – 15	Bi or Tri lobed	Large blue	Few Hours

1.2 Pathology

Pathology is a medical field in which we study the diseases and abnormalities. Pathology word comes from Greek word ‘pathos’ which mean ‘suffering’. History of pathology originated from ancient time. In ancient times, Egyptian started culture to maintain the documents of diseases. At that time’s papyrus they found the information about parasites, cancer, bones injuries and other diseases. In 5_{th} century BC, Hippocrates a Greek physician, profound of pathology made record of diseases and after that many inspired people recorded the detail of diseases such that tumors, tuberculosis, etc. In 19th century, after invention of microscope, a biggest revolution had occurred in pathology. It was first time, when individually cell of an organ deeply studied which help to recognize the diseases. The development of microscope and their availability led to scientific advancement in medical field. Pathology is a broad field, which categories into three major sub types such that anatomical pathology, clinical pathology and molecular pathology. These categories further classified according to study and diversity way of

research of medical.

Figure 1.8 Microscope [3]

1.2.1 Clinical Pathology:

In this field of pathology, disease is diagnosed through laboratory analysis of body's fluid and tissues. For example, analysis the chemical component of blood and cells to identify any bacteria are present in body or not. Clinical pathology also has three sub types such are chemical, hematology and immunology.

For the purpose of sub classification of white blood cells, a hematologist take blood smear and applied preprocessing. After that through microscope, he manually classifies the white blood cells as per knowledge. This manual way is more time consuming and is prone to human error occur. To overcome this problem, a modern and fastest method introduce through modern technology of image processing.

1.3 Background, Scope and Motivation

White blood cells (WBC's) play vital role in human immune system to diagnose diseases and also use as a measurement for healthiness. Sub-classification of white blood cells has great importance as discussed briefly in the above sections. White blood cells make itself different

from other groups of blood by having nucleus and cytoplasm. There are five major sub type of white blood cell: Neutrophil, Lymphocyte, Monocyte, Eosinophil &Basophil. These sub-types classified on the basis of size, color, texture and morphology of nucleus and their cytoplasm.

Traditionally, pathologist or hematologist use microscope to analyze the human blood. Manual analysis process of human blood is more tedious, time consuming and is susceptible to human error. There is a great need of fast, efficient and automatic system. In the last few decades, a progressive advancement came in medical field through medical imaging technology. Most significant of these computational techniques is image processing especially in medical field its outcome is imperious. Pattern recognition and segmentation of imagining technology has become the essential part of each computerized diagnosis system. This innovation of image processing took great revolution in clinical pathology sides. Although there are many automatic techniques introduced which use artificial intelligence and fuzzy logics but still research gaps are there to improve the results and reduce the complexity.

This thesis proposes a solution to sub classification of white blood cells from images by using artificial neural network and image processing techniques. The proposed method automatically classifies the sub types of white blood cells.

1.4 Road Map

This thesis report is organized as:

- Chapter 2 reviews the literature related to techniques of White Blood cell classification.
- Chapter 3 describes the proposed methodology which used us to classification of White blood cell.
- Chapter 4 represents all experiments and results.
- Chapter 5 describe our conclusions and future Works.

Chapter 2 Literature Review

This chapter provides the literature review and basics trends in the field of sub type of white blood cells classification. This chapter organized as:

- Existing Technique
- Comparative Analysis

2.1 Existing Techniques

Many research groups have worked on the same scenario by following different algorithms such that K-Mean clustering [4], water-shed [5], navies Bayes classifier [6], multilayer perceptron [7], Eigen vectors [8], artificial networks and convolution networks [9]. Few of them discussed in this chapter and categorized in the groups on the basis of inputs.

- Texture and Geometrical Features
- Images as Features

2.1.1 Texture and Geometrical Features

Rosyadi et al [4] presented a research in which they used optical microscopic used to produce digital images of blood samples as dataset. They extracted out feature values from each image and process through K-Means clustering for the classification purpose. In this research, sub types of white blood cells were neutrophil, lymphocyte, monocyte, and eosinophil used to classify. At there, process was divided into four phases such that image pre-processing, segmentation, feature extraction and finally classification. First phase, image preprocessing contained the conversion of RGB images to gray scale and binary images. After that in second phase, resizing, cropping and edge detection applied. In feature extraction phase, five geometrical features were considered in this experiment. Those features were normalized area, solidity, eccentricity, circularity and normalized perimeter. Most frequently used K-Means clustering for classification. It works on the basis of Euclidean distance from cluster centroid to particular substance. Shorter distant from centroid of which cluster mentioned that substance classified in that cluster. Furthermore, in this research, their focused was on the significant of each feature with respect to accuracy and effect on accuracy. After their experimentation, they found circularity feature was most significant than other because it got 67 percent accuracy

which was the highest and eccentricity feature had lowest accuracy which was only reached up to 43 percent. So they concluded theirs result on selection of feature more important rather than number of features.

Huang et al [10] produced a research, which introduced a method to segment and recognize the white blood cells. Their research motive was to help hematologist and update the process of white blood cells classifications. They divided their research into number of portions: nucleus segmentation & recognition, feature extraction, and classification. In nucleus segmentation they used leukocyte nucleus enhancer (LNE) in which enhance contrast of nucleus color and suppressed other colors channels. After LNE process, a multiple level of Otsu's thresholding applied to suppress the other than leukocytes cells. In feature extraction, they first used gray level co-occurrence matrix (GLCM) in which 80 textures features extracted per image. After that they also added shape based features such that compactness and roughness. A large dimension's feature set was outcome however they used principal component analysis (PCA) to reduce the dimensions. Classification applied through genetic-based parameter selector (GBPS) in which 50 time crosses validation applied. So their research concluded genetic algorithm more suitable than K-Means clustering.

Gautam et al [6] did their research on the same scenario of leukocyte classification. Their process starts with preprocessing of microscopic images. Preprocessing contained conversion of RGB images to gray scale, contrast stretching and histogram equalization. After that they applied segmentation through Otsu's thresholding, segmented out treated with mathematical morphology operation. In morphology, only closing operation performed which combined form of dilation process followed by erosion. They used only nucleus to extracting geometrical features in which features were perimeter, area, eccentricity and circularity. One the base of these features, Naive Bayes classifier used to classify. Approach of this classifier used estimation of maximum likelihood. This research achieved 80.88 percent maximum accuracy.

S.S. Savkare et al [11] did his research in the segmentation of blood cells and started with preprocessing in which used median filter and Laplacian filter to enhance the utilize information. After preprocessing, it convert the input image to RGB to HSV and HSV to L*a*b color space where L luminosity layer, 'a' chromaticity layer which indicate color fall from red-green axis and 'b' chromaticity layer which indicate color fall blue-yellow axis. K-Mean clustering applied on L*a*b color space for segmentation of blood cells. Furthermore refine to result, it used

morphological operation and watershed algorithm to separate out each cell. The research of segmentation through K-mean clustering produced satisfactory results.

P. Yampri et [8] al also did their research on the same problem of sub classification of white blood cells. They used blood smear images as input and segmented out the white blood cell through color based. Automatic thresholding was applied to segregate nucleus and cytoplasm of cells and those parts cell further processed to extract out features. Eigen cells used to extract features in which followed steps were cell image to vector, computed mean and covariance of vector and then Eigen values and Eigen vectors. Principle of component analysis used to high dimension of Eigen space to lower dimension space. Through their research, they achieved 92% maximum accuracy.

A lot of research has been done for this purpose by using these classification methods. This research has divided the present literature into texture and geometrical features based classification.

Table 2-1 Features Tables

SR No.	Feature Types	Features
1	Geometrical	Perimeter, Area, Eccentricity, Circularity, Solidity, Compactness, Roughness, Eigen Vectors
2	Texture	Gray Level Co-Occurrence Matrix (GLCM), Local Binary Patterns (LBP)

2.1.2 Images as Features

Macawhile et al [12] produce a research, in which worked on white blood cells classifications. In this research they introduced a method to classify WBC's sub types: monocytes, eosinophils, basophils, lymphocytes and neutrophils through microscopic images. Segment out the WBC's by blob analysis of HSV (Hue, Saturation and Value) saturation

components. There used transfer learning technique to classification of WBC's. Transfer learning technique used for pre-trained convolution neural network (CNN) which has ability to classify all other types object. They used three different types of convolution of networks: AlexNet, ResNet-101 and GoogLeNet which were trained more than a millions images. A sample of convolution network architecture is shown below in Figure 2.1.

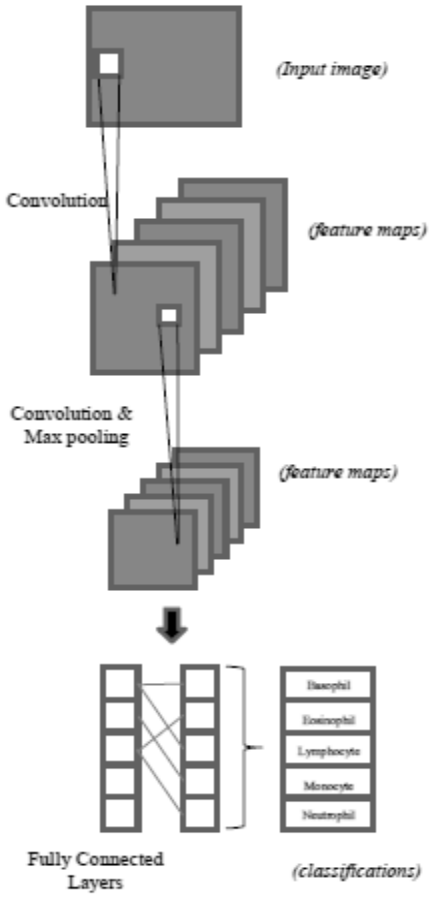


Figure 2.1 Architecture of Convolution Neural Network [7]

Convolution neural network suggest that a huge computation and time required to train. That's why used pretrained CNN which generally used to classify the wide-ranging objects. Although, they were also compared the performance of those pre-trained networks and highly accurate result founded from AlexNet which was 96 %.

Wei Yu et al [13] also did a research on classification of white blood cells (WBC's). They used deep learning method to automatically classify. In deep learning, they also used convolution neural network which was discussed above. They used pre-trained networks such as ResNet500, Inception V3, VGG 16, VGG 19 and Xception. These networks pre-trained through

ImageNet Large Scale Visual Recognition Challenge (1000 classes classified from million images of dataset). Architecture of this research contained three fully-connected layers which shown as below in Figure 2.2.

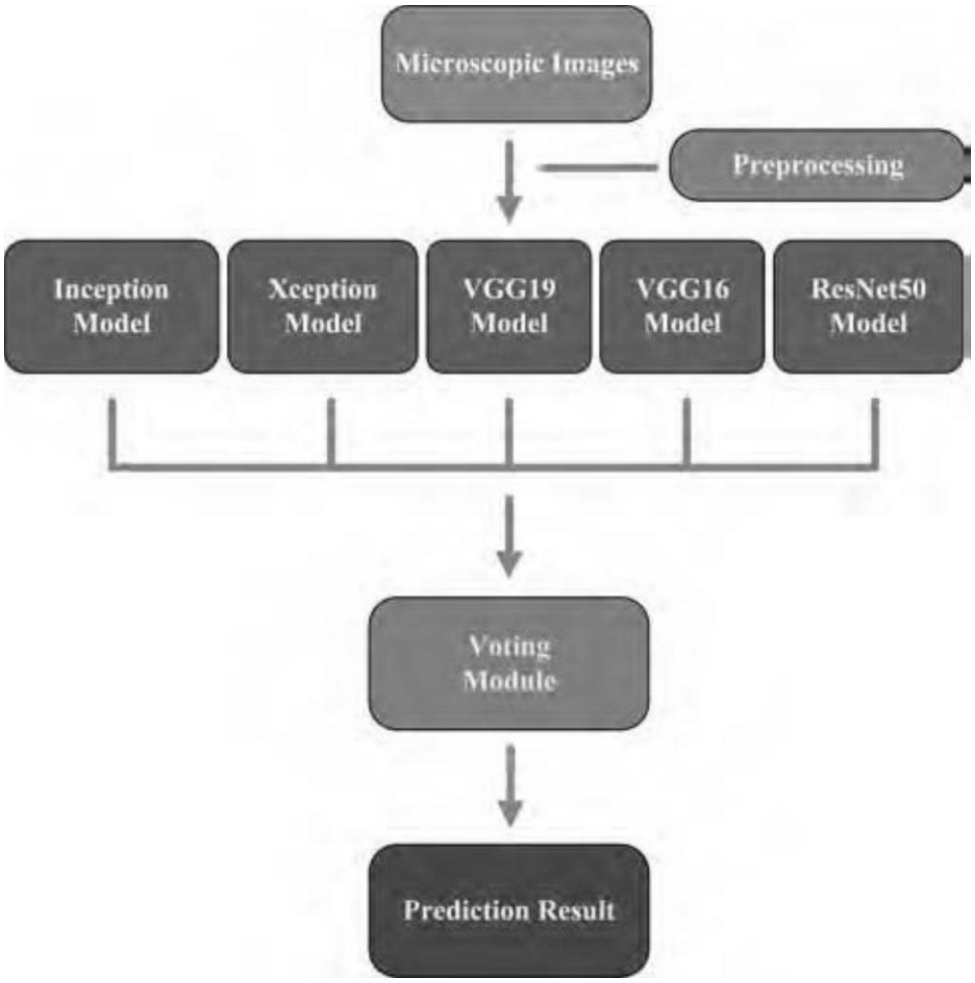


Figure 2.2 Voting Model of CNN's [13]

These models are tuned with respect to requirements and combined their results by voting mechanism to get out final results. After experiment, result of convolution neural network method achieved 88.5 %.

H. Fan et al [14] did a research on white blood cells classification from blood smear images through deep neural network. They proposed a method end-to-end leukocyte localization and segmentation which named 'Leukocyte-Mask'. In this method, they used pixel level prior information to train the deep convolution network as supervise training. The network employed to locate the area of white blood cell and extracted area of the cell forward propagate through neural network to obtained the Leukocyte mask. They validate their experimental results and

made comparison with existing techniques. They claimed their results were more robust and accurate.

O Liang et al [15] did their research on automated blood cells classification through image processing by combining neural networks. Liang et al show a concern about the usage of convolution deep neural network which was incapable to maintain long term dependent relationship between the key features of images and their labels. So they proposed a method which overcome the problem and produced better results. The proposed method was combination of convolution deep neural network (CNN) with recursive neural network (RNN) for deepen understanding of content and features of image. They used transform learning method to pre-train ImageNet dataset for convolution neural network and also used custom loss function for more fast and accuracy. They got nearly 91% highest accuracy through their experimental results which were better than others.

2.2 Comparative Analysis

Table 2-2 Comparison Analysis Chart

Reference	Dataset Standard	No. of Images used Training/Testing	Method	Accuracy
A. Gautaum et al [6]	Yes	20 / 68	Navies Bayes Classifier	80.88%
M. Z. Othman et al [7]	N/A	50	MLP-BP	96%
P. Yampri et al [8]	N/A	50/50	Eigen Vector	92%
Huang et al [10]	N/A	N/A	Genetic Based Parameter Selector GBPS	91%
S.S. Savkare [11]	N/A	70	Water Shed	88.77%
Macawhile et al [12]	NO	21	CCN	96%
Wei Yu et al [13]	NO	N/A	CNN	88.50%
O Liang et al [15]	Blood Cell Kaggle Dataset	12500	CCN + RRN	91%
T. Rosaydi et al [16]	N/A	N/A	K-Mean	67%
S. Manik et al[17]	N/A	90 (54/18)	ANN	95.90%

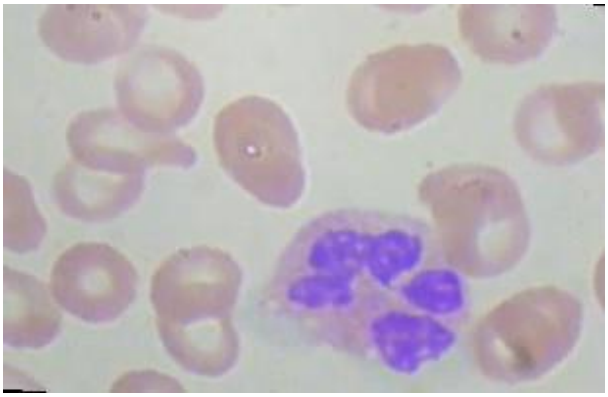
Chapter 3 Proposed Methodology

This chapter provides a detailed overview of the proposed method. The chapter is organized in the following parts,

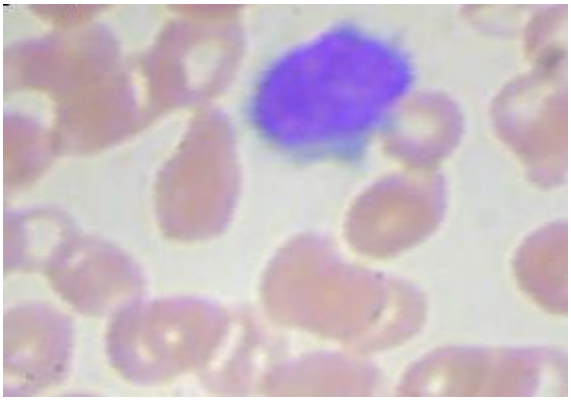
- Dataset
- Development Method
  - Image Pre-Processing & Segmentation
  - Features Extraction
  - Train & Test Classifier

3.1 Dataset

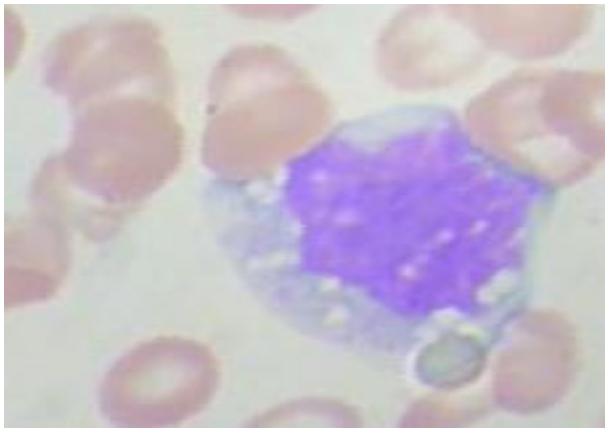
The dataset has been taken from Kaggle which contain 12,500 images of white blood cells (WBC) in JPEG format.



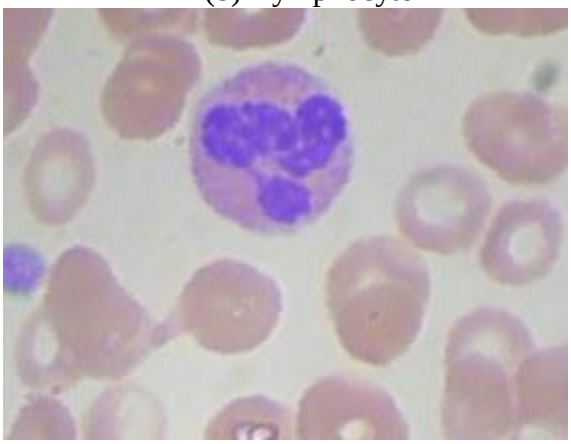
(a) Neutrophil



(b) Lymphocyte



(c) Monocyte



(d) Eosinophil

Figure 3.1 Sub-Types of WBC's

The size of each image is 320*240 pixels. It has four classes, which describe each sub type of white blood cells like Neutrophil, Lymphocyte, Monocyte & Eosinophil can be seen in Figure 3.1. Each class contains approximately 3000 images according to its category. This dataset divided into three portions, i.e. 15%, 15% and 70% ratio. The major part (70%) is used for training the system and other remaining parts are used for validation (15%) and testing (15%) purpose.

3.2 Development Method:

The complete proposed method is divided into the following steps:

1. Image Pre-Processing and Segmentation
2. Features Extraction
3. Classification

3.2.1 Image Pre-Processing and Segmentation

In this step, initially each image of dataset is pre-processed. At this stage, color palette separation is applied to remove background and foreground auxiliary objects. It similar to color based thresholding. All color palette (i.e. red & green) are removed except blue at certain level in order to eliminate the ROI (white blood cell). Furthermore, the image is converted into binary image to perform connected component. After this process, image still has a multiple objects. In order to remove all other axillary objects, area is calculated of required objects and other objects are based on area feature. The segmented cell images are in binary form. Then the exact position of cells through pixel list of binary segmented cell image is extracted. To map the original cell image according to pixel list's segmented cell. Crop the original image and new image size exactly same size of white blood cell (WBC).

3.2.2 Features Extraction

In this step some features, which has useful information based on classification of the sub types of WBC are extracted. For this purpose, this step further divided into three major steps:

1. LBP Features
2. Shape Based Features
3. Hybrid Features

3.2.2.1 LBP Features

Uniform Local binary pattern (LBP) [18] is pattern descriptor techniques which describe an image have to follow the binary pattern in 256 gray levels. It is image operator to transform the image into smaller scale. It used integer labels to present the array or sequence of pixels as metadata. To analysis of the images, used these levels as histogram. LBP feasible all monochrome still images, color images and videos. LBP works on neighbors pixels with respect to center pixel of selective mask. It used threshold, center pixel value define the level. It set the value of corresponding neighbor's pixel to be 1, If the center pixel value is less. Otherwise it sets the concern neighbor to be 0, if it has greater value. It can be expressed as:

$$f(x) = \begin{cases} 1 & if\ \ x \geq n_p \\ 0 & if\ x < n_p \end{cases} \tag{3.1}$$

After thresholding, neighbors' pixels are converted into binary. This process repeats on each pixel of the image, to compute uLBP. The output of each uniform local binary pattern has separate label and non-uniform patterns has separate label.

$$U(LBP_{R,P}) = |s(g_{P-1} - g_C) - s(g_0 - g_C)| + \sum_{p=1}^{P-1} |s(g_P - g_C) - s(g_{P-1} - g_C)|, \tag{3.2}$$

where

p= radius (e.g. 3x3 cell, it is 1).

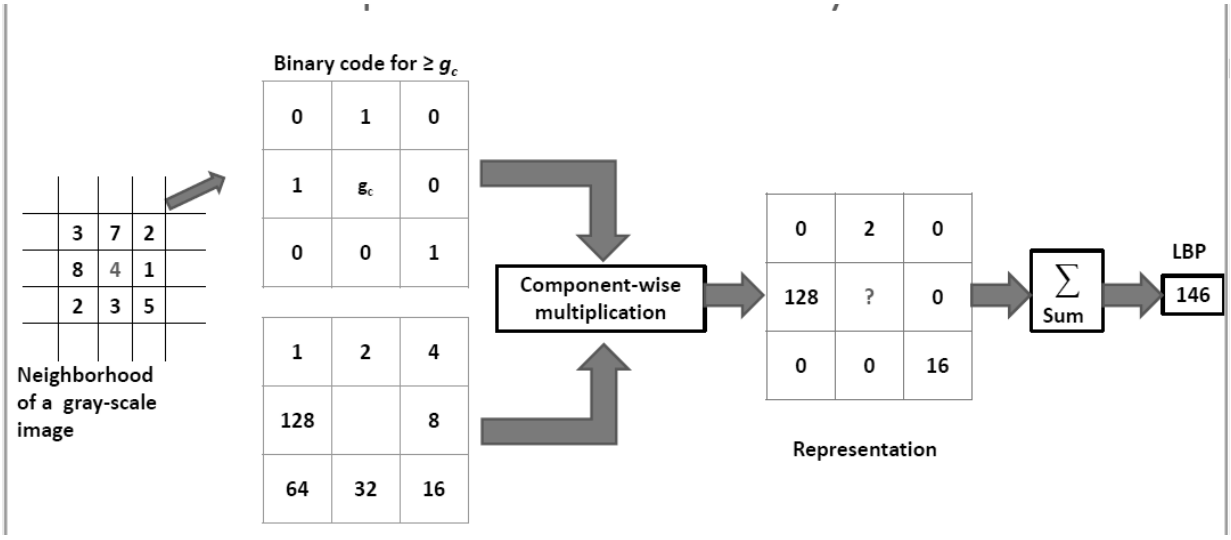


Figure 3.2 LBP Features [18]

3.2.2.2 Shape Based Feature

White blood cell shape well defined architect to distinguish it from its sub types. Each sub types have its' own distinct shape from other in terms of circularity. WBC's types and its diameter are:

Neutrophil (10–12 μm)

Lymphocyte (7–8 μm)

Monocyte(15–30 μm)

Eosinophil (10–12 μm)

After pre-processing and segmenting out the cells, the cells are cropped from the image according to cell and it converted into binary image for calculation of shape based feature. Then we compute shape based feature, which are Major Axis, Minor Axis, Perimeter, Eccentricity and Color Intensity (HSV color space mode). Only these five features worked well to classify but result not reached up to stratification.

3.2.2.3 Combine Features:

For more accuracy and low miss classification rate, combine the both LBP and Shape based feature. It boost the result accuracy to the acceptable level with low cost function.

3.3 Classification

For the purpose of training and classification, we have used two classifiers which are:

- Support Vector Machine (SVM)
- Neural Network (NN).

3.3.1 Support Vector Machine (SVM)

Support vector machine (SVM) [19] is supervised machine learning technique. It is discriminative classifier which defines separable hyper-plane according to each class. It can also be defined, find the optimal hyper-plane in which each class laid either side.

Simply it can be describe as if two classes separately within 2D space, a line can used to separate the classes on the base map data points. But in n-dimensions space, a hyper-plane is used to classify. A dataset define the dimension of a space on the base of number of features. Each feature maps in one dimension space. For n-dimension, hyper-plane equation would be:

$$y = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ \ldots\ldots\ldots$$

$$y = w_0 + \sum_{i=0}^{n} W_iX_i$$

$$y = b_0 + \sum_{i=0}^{n} W_iX_i \tag{3.3}$$

where

W_i=*Vectors* (W_1, W_2, W_3, ...)

b_0=*Biased* (b_0)

X_i= *Variables*(X_1, X_2, X_3, ...)

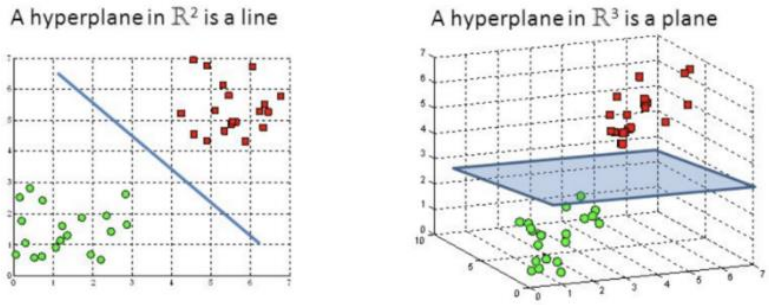


Figure 3.3 Hyper Plane in 2-D & 3-D Space [19]

Furthermore, if data is not properly separable then we move to higher dimension to make separable form. We can assume only up to 3 dimensions. During separation of space with respect to classes, it could not follow simply a line or hyper-plane for good accuracy. A support vector points play important role to define the hyper-plane. Support vector points are those which are closest points of two classes. However, it was difficult to find out the vector points. If founds correctly, it helps to find out best hyper-plane. If the position of vector points are change then hyper-plane is also be changed. Hyper-plane creates at maximum margin of support vectors.

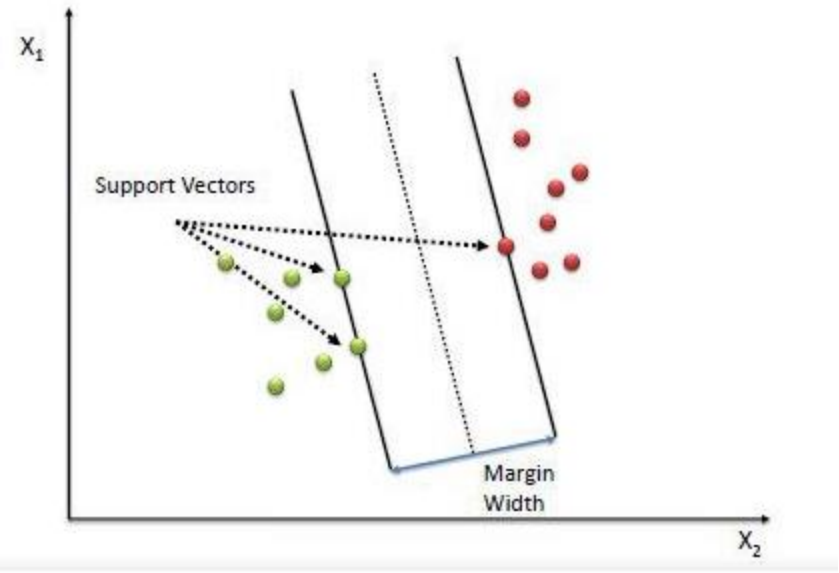


Figure 3.4 Support Vectors [20]

It was more complex whether its soft or hard margin used. This decision depends on the separation of data. If separation of data in a way which can linear classify then hard margin would be used. When used hard margin equation expressed as :

$$\begin{cases} y_i(w^T X_i + b) \geq 1, & Correctly\ classified \\ else, & Misclassified \end{cases} \tag{3.4}$$

Soft margin used when data points are outlier. In this margin category, one slack variable added in the function to skip all outlier data variables. Then function became as:

$$y_i(w^T X_i + b) \geq 1 - \xi_i \qquad \begin{cases} \xi = 0, & Correctly\ classified \\ \xi > 0, & Misclassified \end{cases} \tag{3.5}$$

Slack variable also help to estimation of the error. The value of slack variable express, right dimension or not. The objective has maximum margin from hyper-plane and minimize the error. So it can be express in mathematically as:

$$minimize\ _{W,b} \frac{1}{2}||W||^2 + \sum_{i=1}^{n} \xi_i \tag{3.6}$$

This function describe vector (W) and scalar (b) of maximize the margin and minimize the error to all data points classified correctly. The loss function graphically represented as

Figure 3.5 Loss Function

In real life scenario, data points are not linearly separable. For separation of data points of different classes, it maps to higher dimension and a graphically example given below:

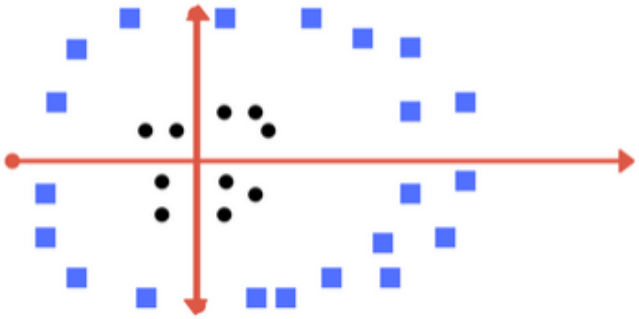


Figure 3.6 Non-Separable Linearly [19]

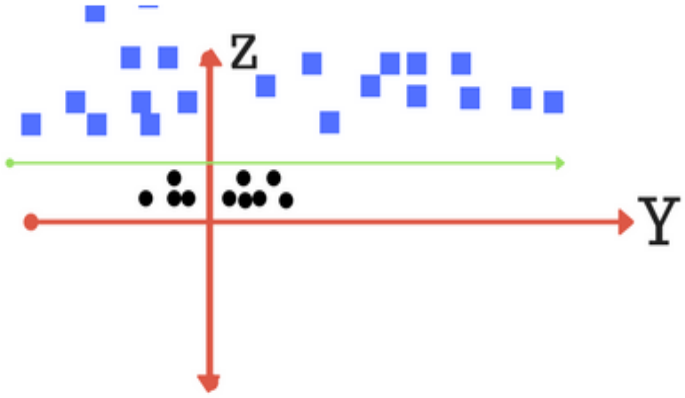


Figure 3.7 Higher Dimension [19]

Maps to higher dimension and draw boundary which makes it separable.

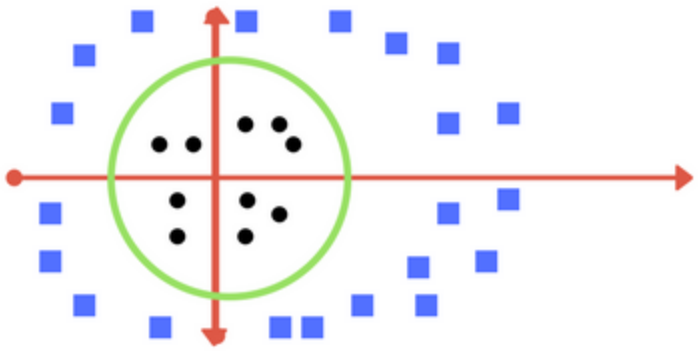


Figure 3.8 Back to Original Dimension [21]

Maps back to original plane. All these transformations done through kernel (Kernel discus later).

So there is need of a support vector machine (SVM), which works as same powerful classifier like two dimensions. Above all formulation discussed, for two dimension space and It also called primal form of SVM. Another SVM form called dual which use Lagrange's multiplier to solve problem in higher dimension.

3.3.1.1 Kernel with SVM

In support vector machine (SVM) kernel is another major part which defined the mode. Kernel of SVM basically a way which computes the dot product between two vectors in high dimensions space (Feature Space). In more generalize, Kernel function can be used as "generalize dot product".

$$maximize_{\alpha} \sum_{i=0}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i^T . x_j) \tag{3.7}$$

where

$$0 \leq \alpha_i \leq C$$
$$For\ all\ i = 1,2,3,\ldots..,n$$

$$\sum_{i=0}^{n} \alpha_i y_i = 0$$

There are number of types of kernels which used to solve the problems. Major types of kernels are:

- Linear
- Polynomial (Cubic & Quadratic)
- Radial Basis function Kernel (RBF) / Gaussian Kernel

- **Linear Kernel**

  Simplest form of kernel which used in simple SVM as discussed above.

- **Polynomial Kernel**

  Polynomial kernel is normally used to compute dot product by increasing the power of kernel function. It can be expressed simple equation as:

$$K(x_1, x_2) = (a + X_1^T X_2)b \tag{3.8}$$

  where

  a = *Constant term*

  b = *Degree of kernels*

  In this research, we have used two polynomial flavor such that cubic and quadratic.

- **Radial Basis function Kernel (RBF) / Gaussian Kernel**

  Gaussian Kernel is another major kernel which used transform the dot product in infinite dimension beyond the higher dimension to Gaussian function. Gaussian function depends on the value of distance from origin point or data point. Mathematically it represent as:

$$\mathrm{K}(\mathrm{X}_1, \mathrm{X}_2) = \mathrm{e}^{-\gamma||\mathrm{X}_1 - \mathrm{X}_2||^2} \tag{3.9}$$

  where

  $x_1 - x_2$ = *Euclidean distance between* x_1 & x_2

  At there, in this research I had used multiple form of Gaussian such that Fine Gaussian, Medium Gaussian and Coarse Gaussian.

3.3.1.2 Tuning Parameters of SVM

Mainly tuning parameter, which control the SVM with respect to kernel is regularization. Regularization parameter can be defined as how to set your model as tight or more generalize. In

polynomial kernel, it describes the value of 'C' as shown in above equation. If the value of 'C' is high, then boundary or hyper-plane will be more complex but the number of misclassification is less. Otherwise if the value 'C' is low then the model will more generalize. The boundary of class will be simple and also misclassification will increase.

For Gaussian kernel, 'ɣ' value defined the model behavior. High value of 'ɣ' model will be tight and less number of misclassification. Low values of 'ɣ', model more generalize but misclassification also increase.

Regularization parameter value had great trade of because if the value is too large then the model will over fit and if the value is too small then model will under fit.

3.3.2 Neural Network (NN)

Neural network is a model which totally inspired from human brain like how neurons connected to each other. A human brain contains approximately 10 billion neurons in which each neuron connect 10000 other neuron. For neural network model, a artificial neuron which known as perceptron which is discovered by 'Frank Rosenblatt'. A simple perceptron takes several inputs and produce single binary output.

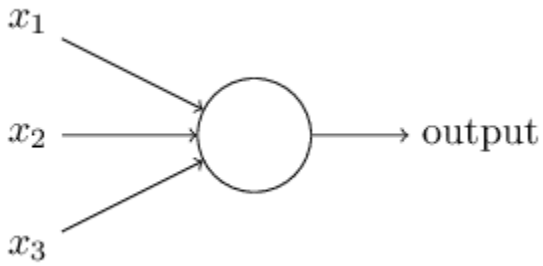


Figure 3.9 Neuron

'Frank Rosenblatt' makes the rule to generate output of perceptron. In this rule, he introduces a weight's factor of real numbers with respect to each input importance. For binary output, sum of all weighted product of input with respect and compression of threshold value. i.e.

$$output = \begin{cases} 0, & \sum_j W_j x_j \leq Threshold \\ 1, & \sum_j W_j x_j > Threshold \end{cases} \quad (3.10)$$

Above equation represent how simple perceptron works but it is very basic model. In real scenario, it may could not work or result will be cumbersome. To overcome this problem, there is need to upgrade. For up-gradation, initially two changes are applied. First changes $\sum_j w_j x_j$ write as dot product of vectors with respect to inputs and weights. Furthermore change, threshold value move to other side of inequality and this term known as bias (bias b = -threshold). Thus

$$output = \begin{cases} 0, & \sum_j W_j . x_j + b \leq Threshold \\ 1, & \sum_j W_j . x_j + b > Threshold \end{cases} \quad (3.11)$$

Bias is another weight vector or neuron which has no input requirement. If bias value is too large then it easy to perceptron output 1 and if it is negative, it difficult to perceptron output to 1.

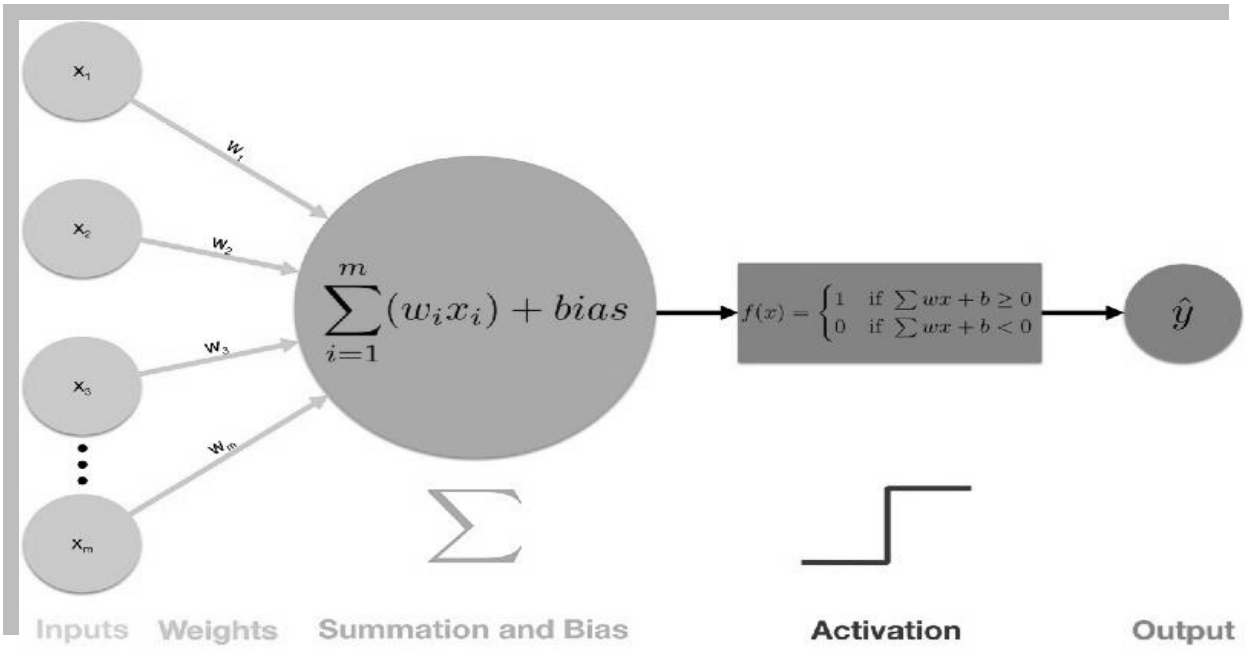


Figure 3.10 Neuron Activation [22]

As shown in above figure, step activation function applied. Some other condition more suitable function may apply like sigmoid, hyperbolic tangent, rectifier and other. A problem had with step function, it can use only for linear function or simple problem. It can't be useful for more complex problem or complex decision. Linear Step function perceptron only use for single layer network, in which a perceptron get input and produce finally output. So, a linear activation function could not effective on multi-layer network because real life problem is not be linear. For incoming nonlinear data, we use nonlinear mapping function called activation function. This

function maps between 0 and 1. Linear function cannot maps between 0 and 1, it maps either 0 or 1. One thing more, linear function cannot reflect small changes in weight while nonlinear function continuous and differentiate able and also maps large range value between 0 and 1. The aim of neural network is produce classified boundary according to activation function.

3.3.2.1 Sigmoid Function

Sigmoid function is monotonic function which has used instead of step function. Similarly, a sigmoid neuron has inputs x_1, x_2, x_3, …, Weights w_1, w_2, w_3,… and bias b. The output of sigmoid neuron is not 0 or 1. It may be any value in the range of 0 to 1 or 1 to -1.

$\sigma\,(w\,.\,x + b)$, where σ is called sigmoid function. Mathematically written as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3.12}$$

Where

$$z = \sum_j W_j . x_j + b$$

Then

$$\sigma(z) = \frac{1}{1 + e^{-(\sum_j W_j . x_j + b)}} \tag{3.13}$$

Sigmoid function graphically represent as:

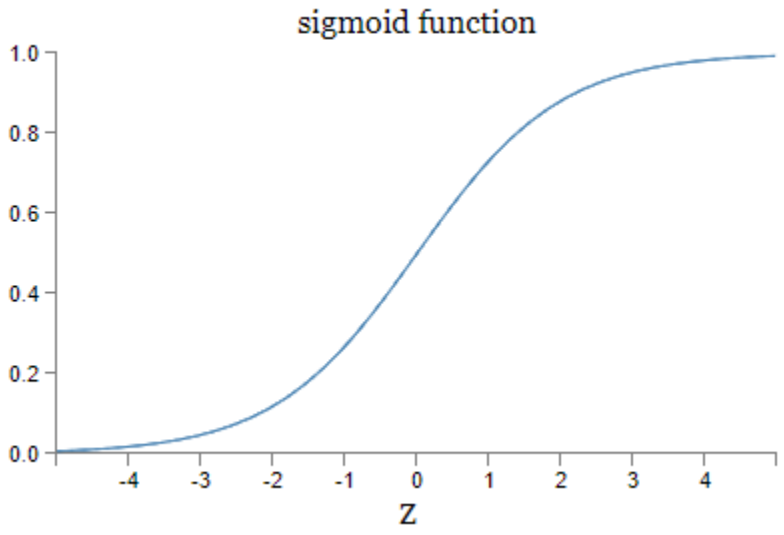


Figure 3.11 Sigmoid Function [22]

After sigmoid function applied, then smoothness of function is clearly visible. Smoothness means if small changes in weights ΔWj and small changes in bias Δb then the output of sigmoid neuron is Δ output (Δ represent change). Sigmoid neuron looks as:

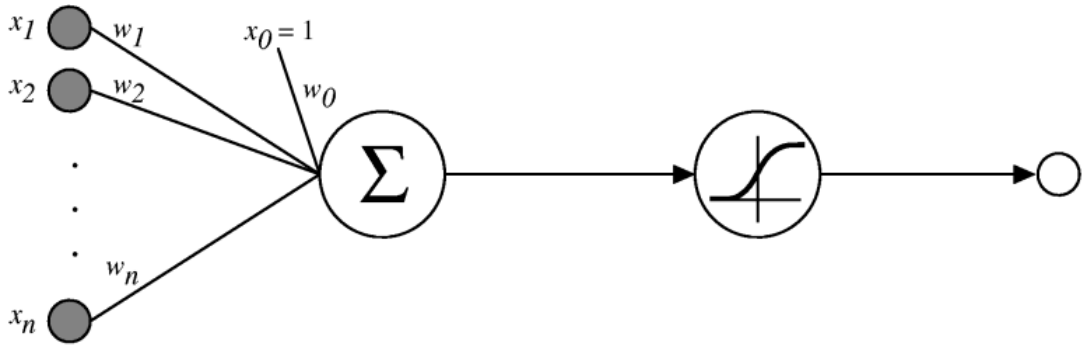


Figure 3.12 Sigmoid Neuron [22]

3.3.2.2 Feed-forward Neural Network

Feed-forward neural network [23] is first type of artificial neural network where the information only travel in forward network, first enter through input layer and then move to hidden layer and finally through output layer. This type of network primarily used for supervised learning. In this networks, sigmoid function used as activation function.

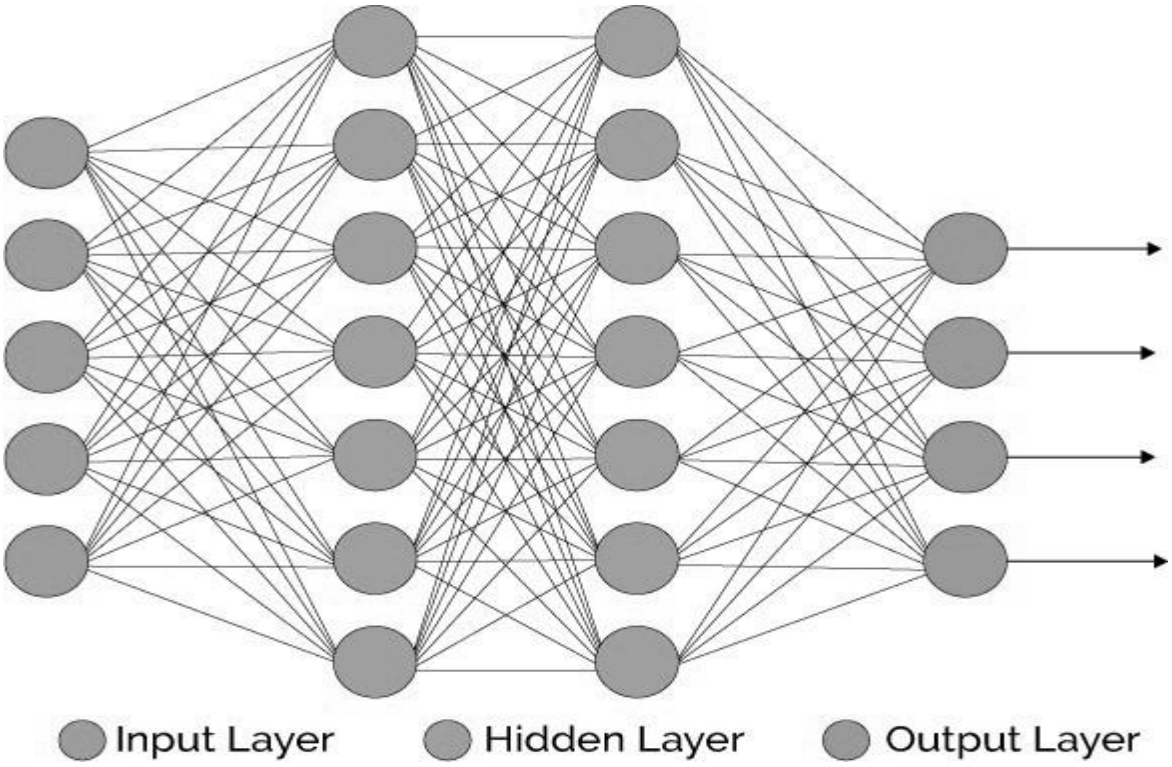


Figure 3.13 Architecture of Feed-forward Neural Network [23]

As show above figure, architecture of feed-forward neural network has divided into three parts:

- **Input Layer**

In this layer, data input at nodes without any processing and its simply feed the data. The output of each node is x_j, where j is going to total number of input d. x_0 is special input, which always is 1. The special input used as bias for hidden layer.

- **Hidden Layer**

Output of the input layer nodes associate with respect to weights w_j as the input of hidden layer nodes. The bias node x_0 has also associate with a weight w_0 and its update similarly as normal weights. But its input value is always is 1. Hidden layer node calculated the summation of weighted output of input layer and applied thresholding (activation) function. Sigmoid used as thresholding function and output of hidden layer node is z_h:

$$zh = \frac{1}{1 + e^{-(\sum_j W_j . x_j + b)}} \tag{3.14}$$

Where

h = 1,2,3,……, (total no hidden nodes)

d = No. of inputs

- **Output layer**

Output layer node nearly computes as like hidden layer node but a slight difference is occurring. The difference depends on problem which used solve this network and also the number of output. At output node, we computed weighted sum of output of hidden layer nodes. At here, we labeled the weight by i notation and hidden nodes with h. Similarly, at this layer bias is also added from hidden layer Z_0 which has input value always be 1. At output node O_i computed weighted sum as:

$$o_i = \sum_{h=0}^{H} V_{ih} Z_{ih} \tag{3.15}$$

Finally, to get output y, the moment of decide which function applied according to problem and number of outputs. If output has only two nodes, then apply simply sigmoid function because easy to decide which class belongs to particular input. Which node has closed value to 1, that type of class assign. If there are more than two nodes, it's very difficult to decide which class classify correctly due to variance between nodes value is not too high. So here to need one more step, a function is required which map one highest value to 1 and others to 0. For this purpose, "Softmax" is used. Softmax function is applied after calculating O_i of each output node. Then y_i as:

$$y_i = softmax(O_i)$$

$$=> \quad y_i = \frac{e^{O_i}}{\sum_{i=1}^{k} e^{O_i}} = \frac{e^{\sum_{h=0}^{H} V_{ih} Z_h}}{\sum_{i=1}^{k} e^{\sum_{h=0}^{H} V_{ih} Z_h}} \tag{3.16}$$

3.3.2.3 Backpropagation

Backpropagation is a method which used to train the network by updating weights. Generally, gradient decent used to update the weights by minimizing the mean square error between actual output and targeted output. Weights updating through partial derivative of error function and each weight tune with respect to its contributions. These weights finely tune after iterative process from output toward input. Mathematically can be expressed as:

$$\Delta V_{ih} = \eta\left(\gamma_i^t - y_i^t\right) Z_h^t$$

$$\Delta W_{hj} = \eta\left(\sum_{i=1}^{k} (\gamma_i^t - y_i^t) V_{ih}\right) Z_h^t (1 - Z_h^t) x_j^t \tag{3.17}$$

Where

X_t = *Input*

Y_t = *Actual Output*

γ_t = *Target Output*

η = *Positive Constant Learning Rate*

t = *No. of Training Dataset*

Chapter 4 Experimental Setup and Results

In this chapter we have discussed implementation of proposed technique and results. This chapter divided into following sections:

- Experimental Setup
- Results
- Comparison

4.1 Experimental Setup

Before start of experimental setup, let's have a little description about machine which used to perform experiment. We have used 7th generation machine which has four CPUs 'Intel core m3' at the rate 1.6 GHz, RAM 8 GB, graphic memory is 4 GB and Windows 10 operating system installed. Matlab used as major tool to computation and perform the experiment.

Table 4-1 System's Description

Generation	7^{th}
Processor	Intel core m3
No. of Processor	4
Speed	1 GHz ~ 1.6 GHz
Memory	8 GB
Graphic Memory	4 GB
Operating System	Windows 10 (64-bit)
Tool	Matlab (R2018a)

After the environment is set, we moved towards dataset. A standard dataset is used to perform the experiment. This dataset obtained from Kaggle which contain 12500 images in JPEG format. Each image has 320 * 240 dimension pixels. The dataset contains four classes: Neutrophil, Lymphocyte, Monocyte and Eosinophil. Dataset is divided into three portion 70%, 15% & 15% with respect to training, validation and testing.

Table 4-2 Dataset Descriptions

Format	JPEG
No. of Images	12500
Dimension	320 * 240
No. of Classes	4
No. of Images in each Class	3000

4.2 Results

This section is further divided into following number of sub section: Segmentation, Feature Extractions, Training and Testing Results

4.2.1 Segmentation:

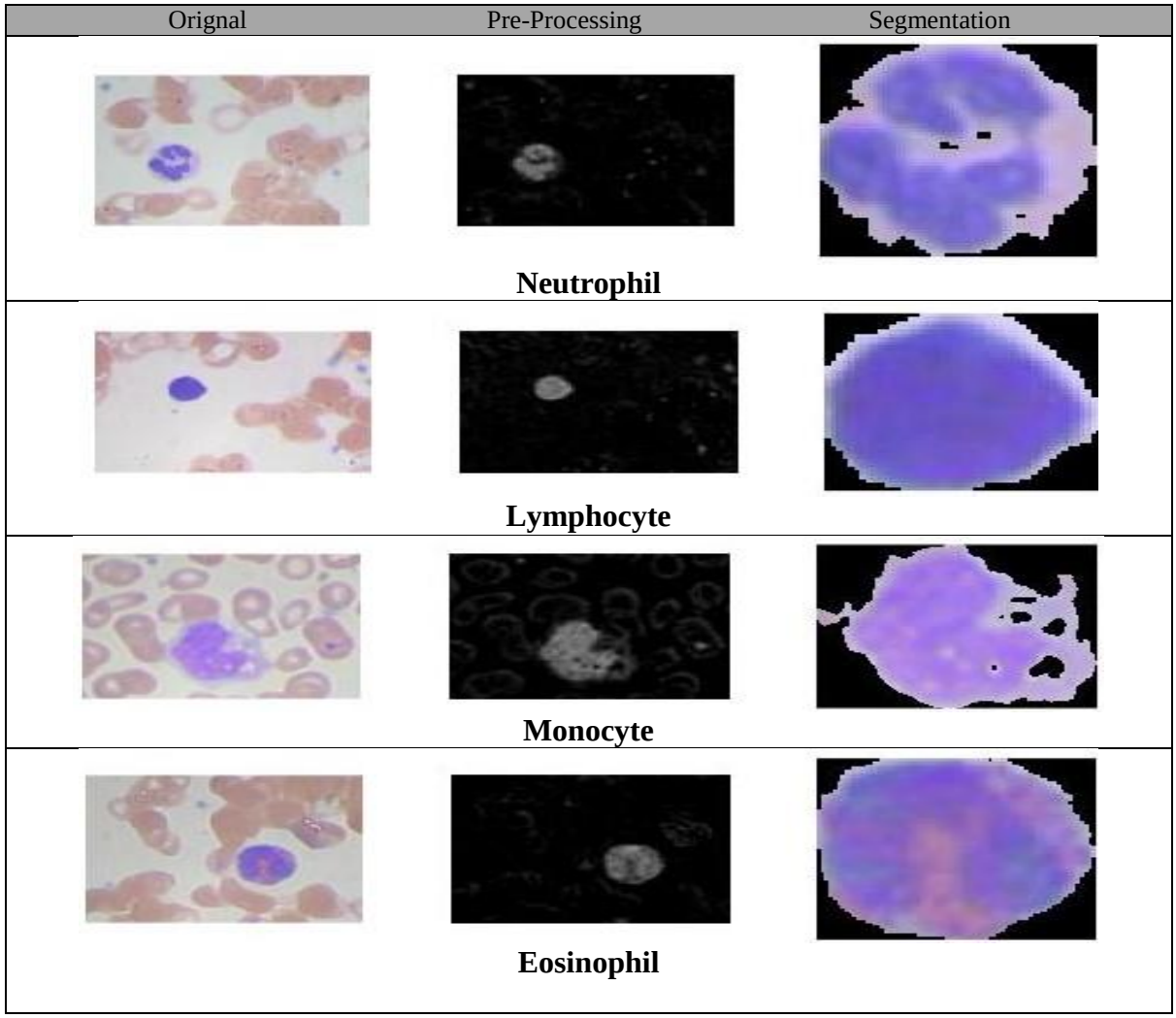


Figure 4.1 Segmentation & WBC Extraction

After preprocessing of images in dataset, segmentation is applied to extract out the white blood cell. Segmentation is applied through color palette separation for removal of background and foreground auxiliary objects. It seems to color base band pass. After cropping, the cell area is extracted out and results are shown in Figure 4.1.

4.2.2 Feature Extraction:

After segmentation and extracting out WBC, features are extracted. Here we used two different ways to extract out:

- Uniform Local binary pattern (uLBP))
- Shape based features

First Uniform Local binary pattern (LBP) features are extract out of each image of dataset and are stored in a matrix. In this matrix, 59 features of LPB are stored. Secondly, Shape based features are extracted out through different computational methods. The four features extracted are Major Axis, Monorails, Perimeter, Eccentricity and Color Intensity. Finally a feature set is prepared by combining both uLBP and shape based features, which we named as hybrid feature set.

4.2.3 Training and Testing Results

Results section has been divided into major two sub sections. These sections are:

- Support Vector Machine (SVM)
- Feed-forward Neural Network

4.2.3.1 Support Vector Machine (SVM)

Support vector machines are used with different kernels and their results are discussed below. For problem of classification, support vector machine (SVM) output represent by confusion matrix. Confusion matrix is a technique which evaluates the performance of a classifier. It shows the count of each class which truly classify or not.

Initially, we experimented with individual feature sets, uniform local binary pattern and shape based feature and their results are shown in Table 4-3. We observed that accuracy

achieved with individual feature sets did not reach at satisfactory level and there is a need to further increase the accuracy..

Table 4-3 Result of Individual Feature Sets

Classifier	Accuracy of LBP Features	Accuracy of Shape Base Features
Linear SVM	73.8%	71.1%
Quadratic SVM	80.50%	75.3%
Cubic SVM	82.90%	69.2%
Fine Gaussian SVM	69.70%	85.7%
Medium Gaussian SVM	83.70%	79.3%
Coarse Gaussian SVM	72.20%	72.8%

After that, we decided to gradually add one shape based feature to LBP feature set to evaluate the effect of each feature. We combined local binary pattern with shape based feature one by one with respect to major axis, minor axis, perimeter, eccentricity and mean color intensity. Classification results with support vector machine (SVM) are shown in Table 4-4.

Table 4-4 Combinational results by adding single feature

Classifier	LBP+ Major Axis (f1)	LBP + f1+Minor Axis (f2)	LBP + f1 + f2+ Perimeter (f3)	LBP + f1 + f2 + f3+ Eccentricity (f4)	LBP + f1 + f2 + f3 + f4+ Mean Intensity
Linear	75.6%	76.1%	77.5%	89.2%	90.3%
Quadratic	85.5%	86.1%	87.4%	94.3%	94.5%
Cubic	86.9%	87.5%	90.0%	94.7%	94.9%
Fine Gaussian	73.7%	73.9%	74.5%	75.4%	75.4%
Medium Gaussian	86.1%	87.1%	88.6%	94.1%	94.3%
Coarse Gaussian	74.4%	76.5%	75.2%	88.3%	88.7%

From above results, we have seen that most powerful feature is eccentricity which boosts the accuracy of each classifier. Mean color intensity feature has least effect on accuracy. So, why these combination have varied accuracy? For this purpose, we have calculated mean and standard deviation of each shape based feature. Their means and standard deviation are shown in the Table-4-5

Table 4-5 Mean and Standard Deviation of Shape based Features

Classes	Major Axis		Minor Axis		Perimeter		Eccentricity		Intensity	
	Mean	Std.(±)	Mean	Std.(±)	Mean	Std.(±)	Mean	Std.(±)	Mean	Std.(±)
EOSINOPHIL	0.084935	0.987084	0.173273	1.033715	0.103933	0.116314	-0.90798	0.730383	0.164443	0.83961
LYMPHOCYTE	-0.93588	0.601113	-0.80195	0.592904	-0.83477	0.388989	-0.80489	0.649215	0.327444	0.967285
MONOCYTE	0.762857	0.793981	0.589779	0.911689	0.697376	0.895364	1.066981	0.225516	-0.32159	0.96914
NEUTROPHIL	0.088577	0.751623	0.038854	0.853565	0.133977	0.775943	0.549061	0.505855	-0.17077	1.07464

Mean and standard deviation defines specific range of each feature with respect to each class. So we can see that in Table 4-5, each class has its own range but some classes have very close mean values for some features and it is difficult to separate the classes if they have closer means. Over all, eccentricity and major axis have shown proficient ranges of each class as compared to other features.

- **Linear SVM**

Through linear SVM kernel, overall accuracy reached up to 90.3 %. Confusion matrix of all classes has shown in Figure 4.2.

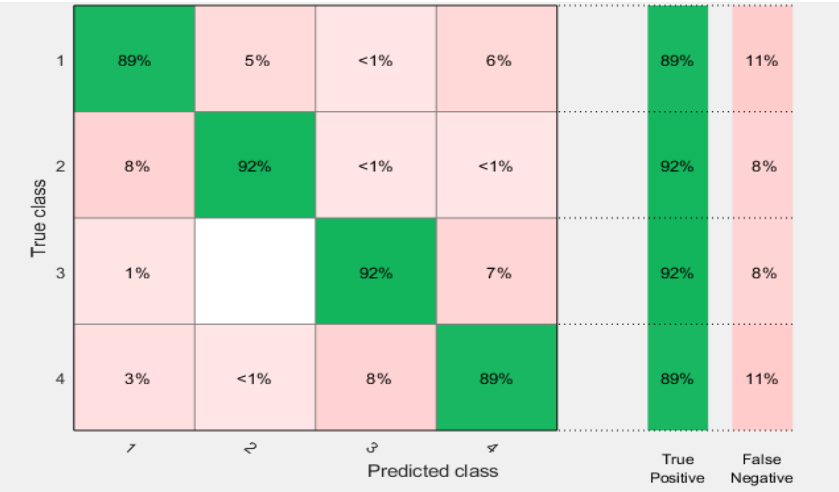


Figure 4.2 Confusion Matrix of Linear SVM

- **Quadratic SVM:**

Through quadratic SVM kernel, overall accuracy reached up to 94.5 %. Confusion matrix of all classes has shown in Figure 4.3.

Figure 4.3 Confusion Matrix of Quadratic SVM

- **Cubic SVM:**

Through cubic SVM kernel, overall accuracy reached up to 94.9 %. Which is the highest accuracy achieved with SVM classifier. Confusion matrix of all classes has shown in Figure 4.4.

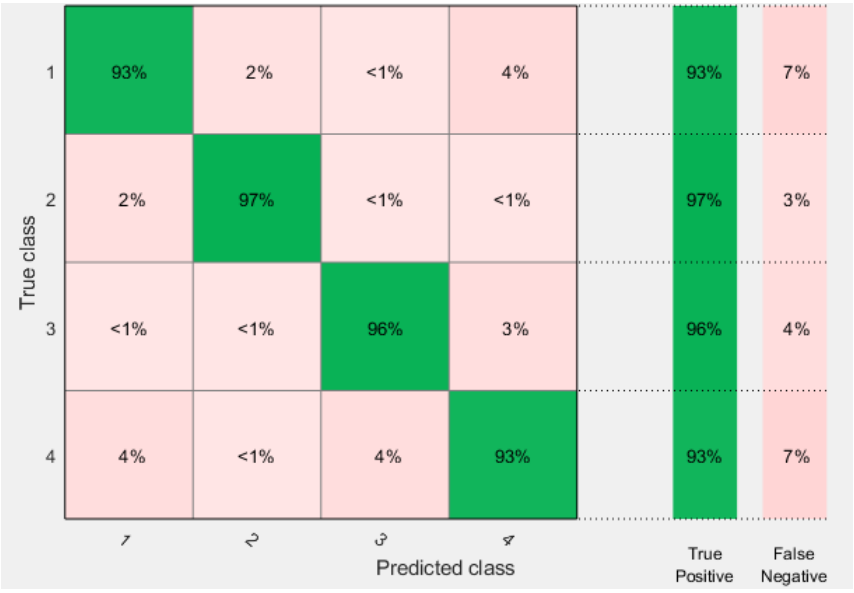


Figure 4.4 Confusion Matrix of Cubic SVM

- **Fine Gaussian SVM:**

Through fine Gaussian SVM kernel, overall accuracy reached up to 62.8 % and it shown lowest accuracy in this case. Confusion matrix of all classes has shown in Figure 4.5.

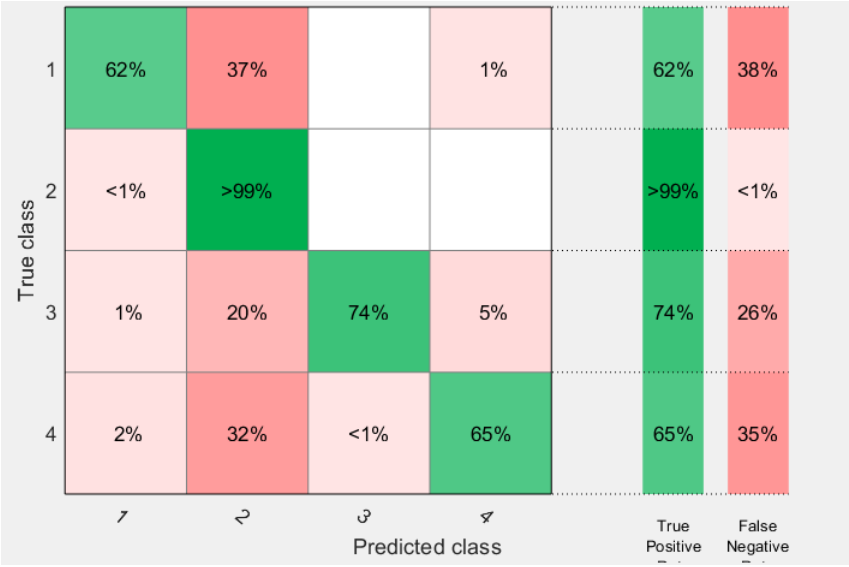


Figure 4.5 Confusion Matrix of Fine Gaussian

- **Medium Gaussian SVM**

Through medium Gaussian SVM kernel, overall accuracy reached up to 94.3 %. Confusion matrix of all classes has shown in Figure 4.6.

Figure 4.6 Confusion Matrix of Medium Gaussian SVM

- **Coarse Gaussian SVM**

Through medium Gaussian SVM kernel, overall accuracy reached up to 88.7%. Confusion matrix of all classes has shown in Figure 4.7.

Figure 4.7 Confusion Matrix of Coarse Gaussian SVM

From these confusion matrices, we can see that Medium Gaussian kernel has shown highest accuracy while Coarse Gaussian has lowest with support vector machine (SVM) classifier. Medium Gaussian has intermediate scaling between two classes, which used scaling set of square root of predictor. While Coarse Gaussian used short scaling set between the classes. In our case, data is not highly distinct because each cells have nearly same architecture and a little detail is distinct from each other. That is why Medium Gaussian worked better than other SVM kernels. We conclude all results of different kernels of SVM classifier in Table 4-3.

4.2.3.2 Neural Network

Neural network achieved best performance with three layered network in which only one hidden layer used with 150 neurons. The performance network measure through cross entropy loss function which is widely used to performance of neural network. Cross entropy fiction value

increase if predicated value and actual value is not same. In ideal case, cross entropy value will be zero.

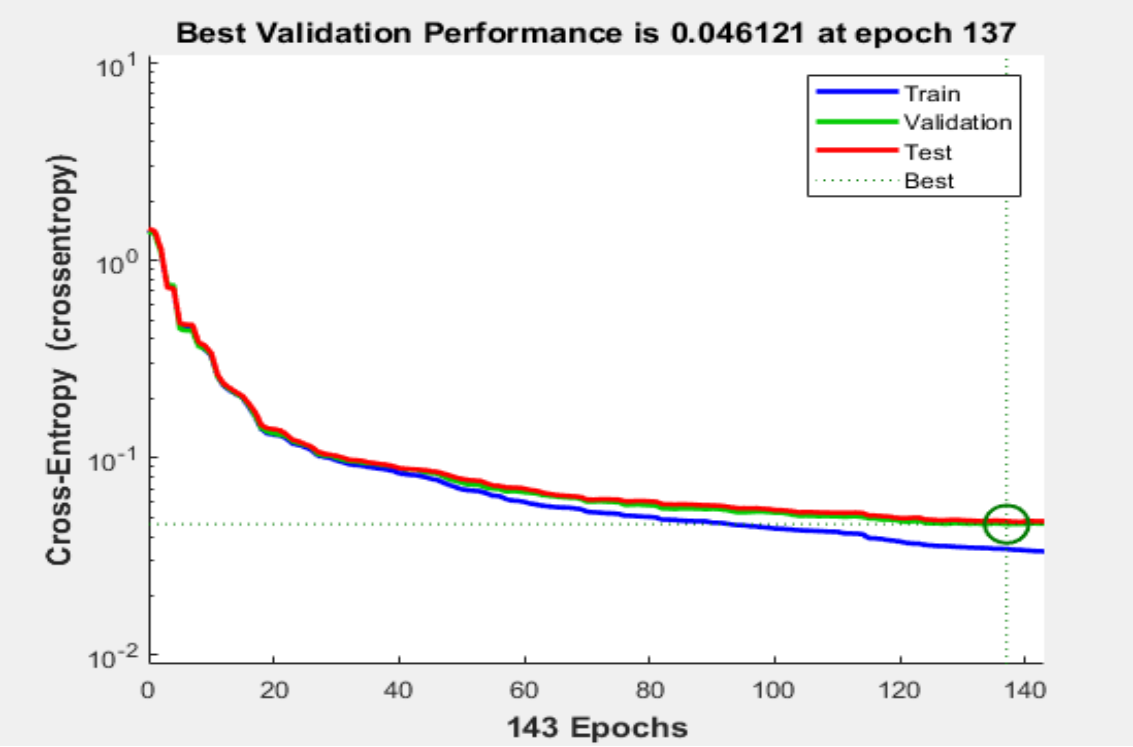


Figure 4.8 Cross-Entropy

The cross entropy of the neural network shown in Figure 4.8 with respect to training, validation and testing. In this case, each epoch trained the network and calculated cross entropy. This process stopped when it reached at minimum level and the value would constant on further each epoch. In our instance, the minimum value of cross entropy is 0.046121 after 137 epochs. It seemed to be nearby zero.

Error histogram is another way to evaluate the performance of neural network. In which error classified into bins and each bin has discrete range error. Each bin of histogram presents a range of error. More weighted histogram means more probably occurred. Error histogram has shown those occurred error during the training of network.

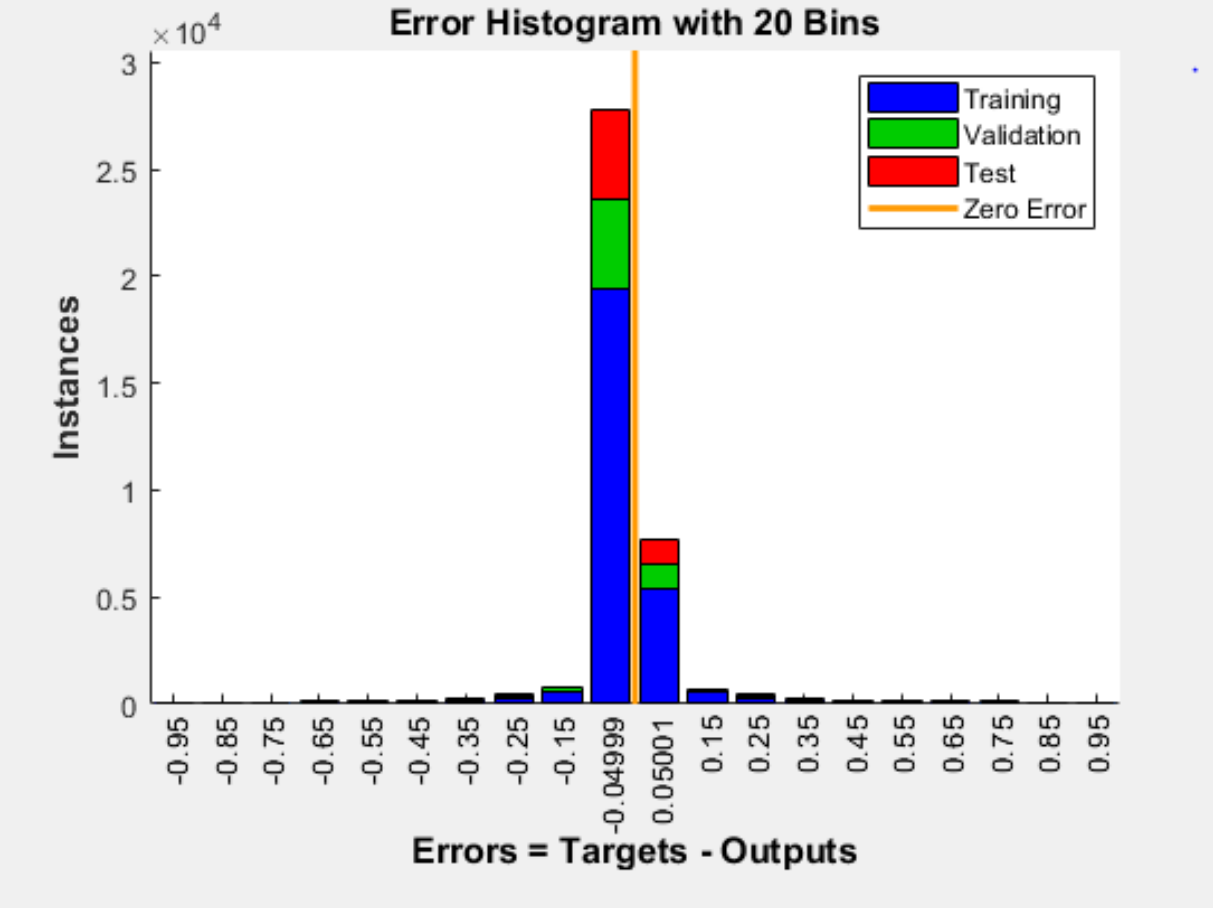


Figure 4.9 Error Histogram

In our case maximum error seemed to be 0.04999 which is combined form of training, validating and testing. As shown in the Figure 4.9, blue, green and red colors represent the error with respect to training, validating and testing. Note that error is normalized form in the range of 1 to -1.

For output of neural network to classification problem is confusion matrix. Confusion matrix basically a technique which summarize the performance of a classifier. The main idea to fill the confusion by passing data to classifier and it predicate result match with the original, find count with respect to true and false classes. The output of the network shown in Figure (4.10, 4.11 and 4.12) respectively training, validating and testing.

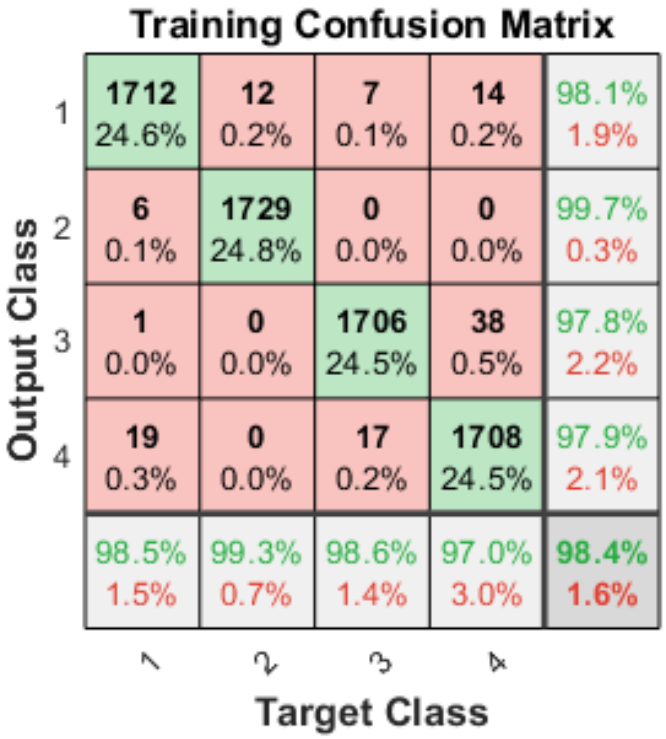


Figure 4.10 Confusion Matrix of Training Dataset

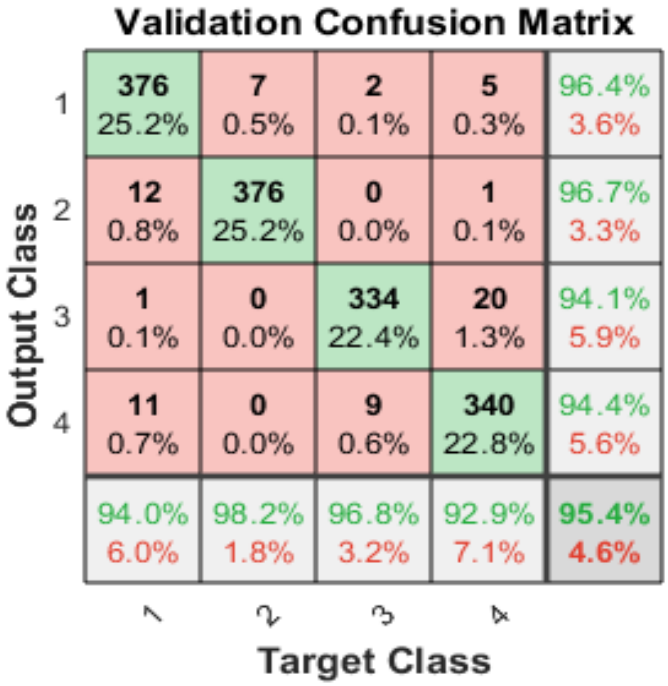


Figure 4.11 Confusion Matrix of Validation

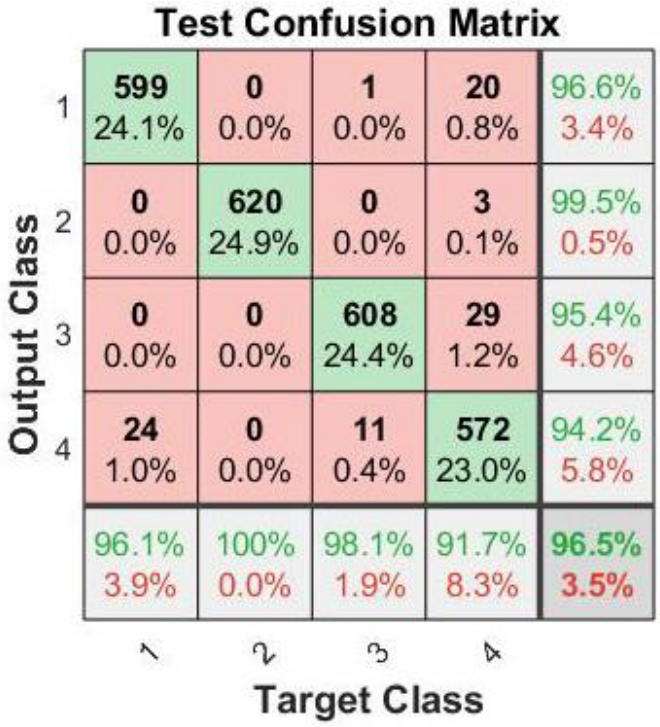


Figure 4.12 Confusion Matrix of Test Dataset

In confusion matrix, 'output class' is a class which classifier's predict and 'target class' is actual's class which it belong. As shown in the Figure 4.12, we have seen that classifier had mostly confused class 1 (Eosinophil) and class 4 (Neutrophil). Because class 1 (Eosinophil) and class 4 (Neutrophil) have almost same size of diameter and multi-lobed nucleus, so feature values related to them also come out almost similar. Overall, our classifier got maximum accuracy of 96.5%.

While another best way to evaluate the classifier through receiver operating characteristic (ROC) curve graph. ROC curve draw through only two parameters:

- **True Positive Rate (TPR):**

  This parameter calculated as:

$$TPR = \frac{TP}{TP + FN} \tag{4.1}$$

- **False Positive Rate (FPR):**

  This parameter calculated as:

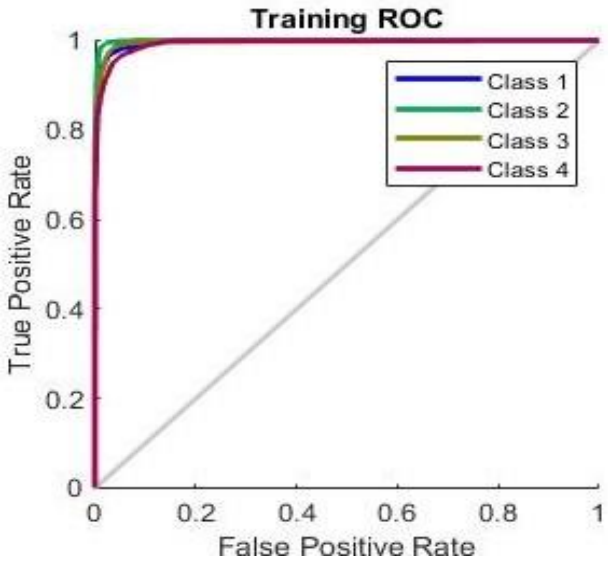
$$FPR = \frac{FP}{FP + TN} \tag{4.2}$$

Where

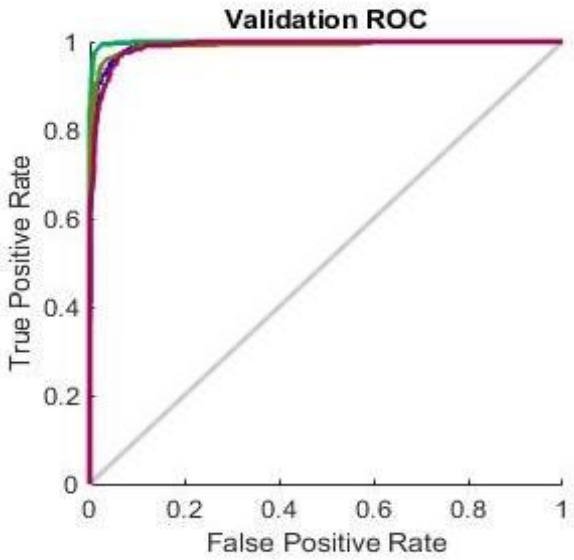
TP = True Positive Class FP = False Positive Class

TN = True Negative Class FN = False Negative Class

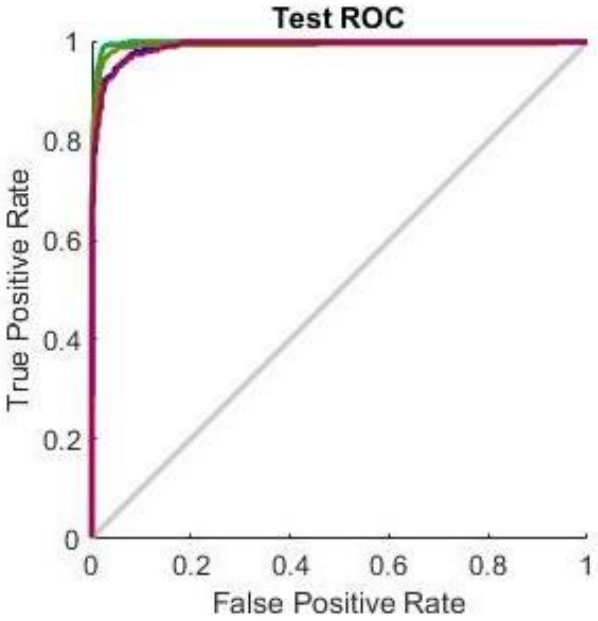
Receiver operating characteristic (ROC) curves graph of our neural network as shown four curves with different colors. Each color of curve represents a class perdition rate as discussed above. Three graphs are shown in Figure 4.13 with respect to training, validating and testing.



a) Training



b) Validation



c) Testing

Figure 4.13 ROC Graphs

To evaluate the ROC graph in multiclass, area under curve (AUC) is used to represent the performance. Area under curve near to 1 represents good classifier and else vice versa. Here, we can see that our model's area under curves of multiple class is near to 1. ROC graph is shown in Figure 4.13 represents good performance of the model.

Highest accuracy of 97.5% has been achieved with neural network. After that we have validated the results by averaging them over ten runs, hence average maximum accuracy achieved is 96.5%

4.3 Comparison with Previous Work

Here we have compared our results of proposed method with other related works. The comparison is shown in Table 4-4. As we can see that our model's accuracy is competitive as compared to previous methods. Studies in [6-8, 17] did not use benchmark datasets and designed their own datasets. They used different types of classifiers as well as different methodology for the purpose of classification of white blood cell (WBC).

Table 4-6 Comparison with Related Works

Reference	Same Dataset	No. of Images	Method	Accuracy
A. Gautaum et al [6]	No	20 / 68	Navies Bayes Classifier	80.88%
M. Z. Othman et al[7]	No	50	MLP-BP	96%
P. Yampri et al [8]	No	50/50	Eigen Vector	92%
Huang et al [10]	N/A	N/A	Genetic Based Parameter Selector GBPS	91%
S.S. Savkare [11]	No	70	Water Shed	88.77%
Macawhile et al [12]	No	12000	CNN	96%
Wei Yu et al [13]	N/A	N/A	CNN	88.50%
O Liang et al [15]	Blood Cell Kaggle Dataset	12000	CCN + RRN	91%
T. Rosaydi et al [16]	N/A	N/A	K-Mean	67%
S. Manik et al[17]	N0	90 (54/18)	ANN	95.90%
Qian Wang et al[24]	No	N/A	SVM	94%
Proposed Method	Blood Cell Kaggle Dataset	12000	Neural Network	96.5%

O Liang [15] used same dataset from Kaggle as we have used in our experimental methodology. O Liang [15] divided the dataset into two portions of training and testing dataset with ratio of 80 and 20 percent. They used convolution network with recursive neural network for classification purpose. They achieved 91% accuracy after complex and deep computations. We have introduced most simple artificial neural network with single hidden layer. Through proposed mechanism, we have achieved average accuracy of 96.5 % after 10 iterations with training and testing division of 80% and 20%, which highlights that our proposed method is competitive as compared to previous methods in literature.

Chapter 5 Conclusion and Future Work

5.1 Conclusion

In this research a novel approach is used to classify the sub types of white blood cells (WBCs) i.e. Monocytes, lymphocytes, eosinophils and neutrophils are using image processing and machine learning. This approach uses neural network and Support vector machine (SVM) for classification and optimize the result of both classifiers. The approach is divided into three parts i.e. preprocessing & segmentation, feature extraction and classification. A microscopic image of dataset containing a large number of different type of blood cells are segmented out during the preprocessing & segmentation part. Color base segmentation applied because only white blood cell has distinct color cytoplasm than other blood cells. After segmentation, feature extraction process starts. Sub types of white blood cells are distinguishable by the physical architecture such as nucleus lobs, cytoplasm, size and shape of cells. Feature extraction process extracted out two type of feature i.e. Shape based feature (Major Axis, Monorails, Perimeter, Eccentricity and Mean color intensity) and uniform local binary pattern (LBP). Then both type of features are combined to make a hybrid set of features. This feature set is fed into support vector machine (SVM) and three-layer feed forward neural network side by side. These classifiers are trained through large data set of blood cell images.

The proposed method is tested on high resolution microscopic images of blood cells dataset. During testing, the proposed algorithm has shown optimal performance in term of classification. The achieved accuracy through SVM is 94.5 % and neural network is 96.5%. So here we observed neural network has produced more optimal results. It also observed that proposed method is fully automatic and edge over on some previous method which are more complex and high resource required to perform computation to produce a certain level of accuracy. But this method is simpler than those and given more optimum accuracy.

5.2 Future Work

The proposed technique may be extended to other tissues or cells of body which may help to biomedical diagnosis. Almost the same algorithm with some modification may be tested for other blood cells like red blood cells, platelets or other tissues.

Other possible approaches can be explored in future which include an implementation of algorithm on neural network may more accurate than proposed method.

REFERENCES

[1] HealthLinkBC, “Components of the Immune System.” [Online]. Available: https://www.healthlinkbc.ca/health-topics/tp10165.

[2] O. Ryabchykov *et al.*, “Leukocyte subtypes classification by means of image processing,” in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2016, pp. 309–316.

[3] “Microscope Images.” [Online]. Available: https://www.shutterstock.com/search/microscope.

[4] N. Theera-Umpon and S. Dhompongsa, “Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, pp. 353–359, 2007.

[5] L. B. Dorini, R. Minetto, and N. J. Leite, “Semiautomatic white blood cell segmentation based on multiscale analysis,” *IEEE J. Biomed. Heal. informatics*, vol. 17, pp. 250–256, 2012.

[6] A. Gautam, P. Singh, B. Raman, and H. Bhadauria, “Automatic classification of leukocytes using morphological features and naïve Bayes classifier,” in *2016 IEEE Region 10 Conference (TENCON)*, 2016, pp. 1023–1027.

[7] M. Z. Othman, T. S. Mohammed, and A. B. Ali, “Neural network classification of white blood cell using microscopic images,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, 2017.

[8] P. Yampri, C. Pintavirooj, S. Daochai, and S. Teartulakarn, “White blood cell classification based on the combination of eigen cell and parametric feature detection,” in *2006 1ST IEEE conference on industrial electronics and applications*, 2006, pp. 1–4.

[9] M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, “Microscopic Blood Cell Classification Using Inception Recurrent Residual Convolutional Neural Networks,” in *NAECON 2018-IEEE National Aerospace and Electronics Conference*, 2018, pp. 222–227.

[10] D.-C. Huang and K.-D. Hung, “Leukocyte nucleus segmentation and recognition in color blood-smear images,” in *2012 IEEE International Instrumentation and Measurement Technology*

Conference Proceedings, 2012, pp. 171–176.

[11] S. S. Savkare and S. P. Narote, “Blood cell segmentation from microscopic blood images,” in *Proceedings - IEEE International Conference on Information Processing, ICIP 2015*, 2016.

[12] M. J. Macawile, V. V. Quiñones, A. Ballado, J. Dela Cruz, and M. V. Caya, “White blood cell classification and counting using convolutional neural network,” in *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, 2018, pp. 259–263.

[13] W. Yu *et al.*, “Automatic classification of leukocytes using deep neural network,” in *2017 IEEE 12th International Conference on ASIC (ASICON)*, 2017, pp. 1041–1044.

[14] H. Fan, F. Zhang, L. Xi, Z. Li, G. Liu, and Y. Xu, “LeukocyteMask: An automated localization and segmentation method for leukocyte in blood smear images using deep neural networks,” *J. Biophotonics*, pp. e201800488–e201800488, 2019.

[15] G. Liang, H. Hong, W. Xie, and L. Zheng, “Combining convolutional neural network with recursive neural network for blood cell image classification,” *IEEE Access*, vol. 6, pp. 36188–36197, 2018.

[16] T. Rosyadi, A. Arif, B. Achmad, and others, “Classification of leukocyte images using k-means clustering based on geometry features,” in *2016 6th International Annual Engineering Seminar (InAES)*, 2016, pp. 245–249.

[17] S. Manik, L. M. Saini, and N. Vadera, “Counting and classification of white blood cell using artificial neural network (ANN),” in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 2016, pp. 1–5.

[18] A. A. Shirazi, A. Dehghani, H. Farsi, and M. Yazdi, “Persian logo recognition using local binary patterns,” in *3rd International Conference on Pattern Analysis and Image Analysis, IPRIA 2017*, 2017.

[19] S. Patel, “SVM (Support Vector Machine).” [Online]. Available: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72.

[20] “Support Vector Machine - Classification (SVM).” [Online]. Available: https://www.saedsayad.com/support_vector_machine.htm.

[21] J. Laosai and K. Chamnongthai, “Acute leukemia classification by using SVM and K-Means clustering,” in *2014 International Electrical Engineering Congress (iEECON)*, 2014, pp. 1–4.

[22] A. Foucquier, “Scheme of one neuron layer with the application of the activation function.” [Online]. Available: https://www.researchgate.net/figure/Scheme-of-one-neuron-layer-with-the-application-of-the-activation-function_fig5_257548537.

[23] R. Bhatia, “Neural Networks Architacture layers.” [Online]. Available: https://medium.com/datadriveninvestor/when-not-to-use-neural-networks-89fb50622429.