

CLASSIFICATION AND SEGMENTATION OF LEUKOCYTES IN MICROSCOPIC BLOOD SMEAR IMAGES USING K-MEANS AND SVM

¹Dr.C.Balasubramanian, ²J.Veneeswari, ³R.Kalaivani

¹Professor,^{2,3}Assistant Professor

¹Computer Science and Engineering,

¹P.S.R Engineering College, sivakasi, Tamilnadu,India.

Abstract : Human body is effectively tested by the WBCs count in the blood smear images. Based on the counting hematic diseases can be found. The hematic diseases are AIDS, blood cancer etc., Here a computer aided blood service analysis for detection and counting of leukocytes in blood samples. White blood cells are otherwise called as leukocytes or leucocytes. They help to protecting body against the infection diseases. By implementing the effective method for counting WBCs. The analysis of leukocytes information is valuable to the medical specialists that help to identify the hematic diseases. In the proposed framework, the microscopic blood smear image nucleus is first segmented, followed by extraction of statistical, texture and geometric features. The image classification is done by multi class classifier (SVM). Finally, the detected leukocytes are classified into five types: basophils, eosinophil, neutrophil, lymphocyte, and monocyte. This method will achieve the good experimental performance.

IndexTerms - Hematology, Image Classification, Segmentation, Leukocytes Classification, k-Means & SVM.

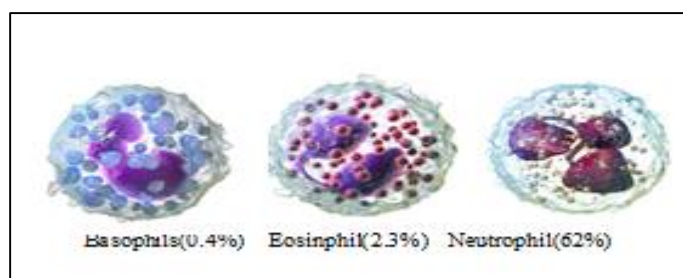
I. INTRODUCTION

White blood cells of human body containing 70,000 cells per microlitre of blood. The number of white blood cells increases; it is a sign of an infection somewhere in the body [1]. The blood is made up of liquid and solids. The liquid part is called plasma. The plasma part contains water, salts, and protein. Over half of blood is plasma. The solid part of blood contains red blood cells, White blood cells and platelets. Red blood cells are the most common type of blood cell [2]. It also called as erythrocytes, deliver oxygen from lungs to tissues and organs. Blood cells constantly die and body makes new ones. Red blood cells live about 120 days. The cytoplasm of erythrocytes is rich in hemoglobin, an iron-containing biomolecule that can bind oxygen and is responsible for the red color of the cells. White blood cells (WBCs) otherwise called as leukocytes or leucocytes [3]-[4]. Leukocytes are component of immune system that is used to protecting the body against infectious disease. WBCs are categorized into five classes: basophils, eosinophil, neutrophil, lymphocyte, Monocyte.

The WBC of each type can be determined with a blood test is called differential count. Basophils are type of immune cell that is responsible for infections, inflammatory reactions, and asthma. 0.4% of basophils present in leukocytes. In WBCs, eosinophil present 5000 cells per microlitre of blood. It is answerable for an extreme allergic reaction, [parasitic infections](#) and drug infections. A neutrophil is the one of the important part of leukocytes. 62% of neutrophils present in WBCs. It is liable for bacterial and fungal infections. These three classes of WBCs are called as granulocytes. They can be characterizing by the existence of [granules](#) in their [cytoplasm](#). The other two types of leukocytes are called as Agranulocytes [5]. It is also known as mononuclear leukocytes. Lymphocyte contains three types of cells. The cells are natural killer cell, B-cell, T-cell. WBCs contain 30% of lymphocyte. Monocyte is the subtype of leukocytes. The lifetime of monocytes is much longer lived cells. 5.3% of monocyte present in WBCs. The following steps to detect and counting the WBCs in blood smear images.

1. Pre-processing
2. Segmentation
3. Feature extraction
4. Classification

Let us see these steps in detail in the following sections:



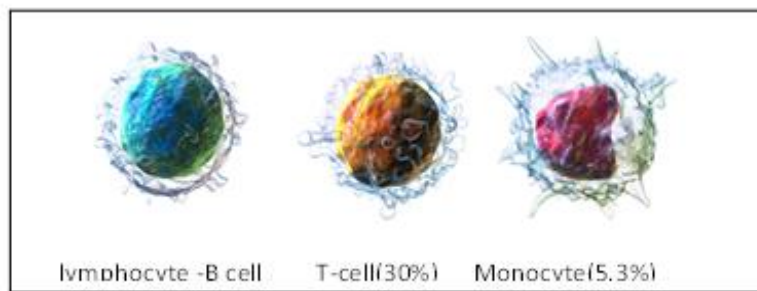


Fig.1. Classification of cells

II. PROPOSED SYSTEM

The proposed system will clearly clarify the proposed system in the graphical way.

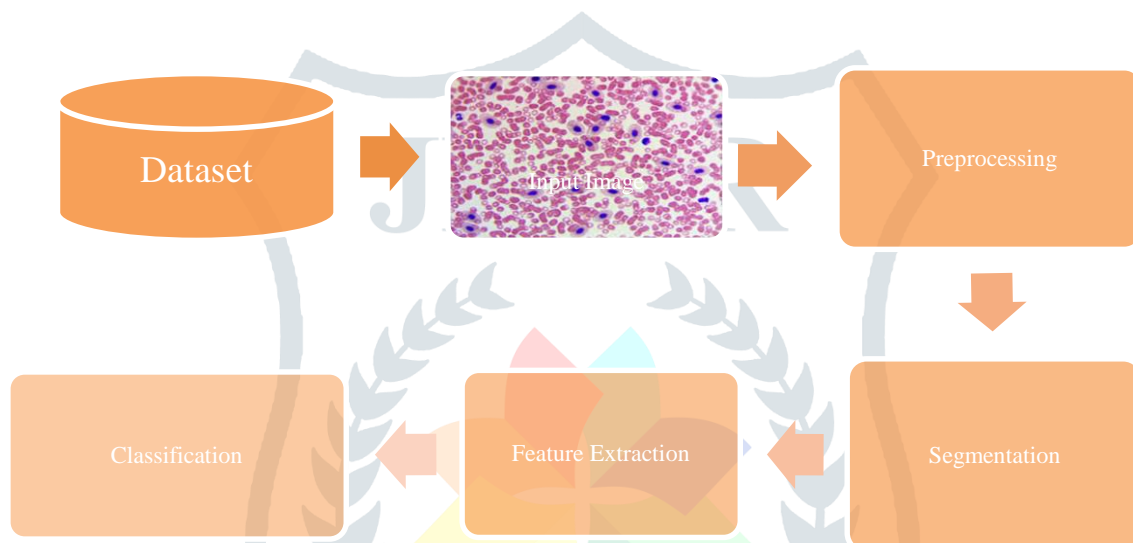


Fig.2. System diagram for proposed method

From the block diagram, first the dataset of WBCs samples are collected from various hospitals. The input images are taken from the dataset. The format of given input image can be in jpg format [3]-[7]. During processing of MATLAB, the size of the image can be adjusted to 256*256 pixels.

The following section clearly explains these modules in detail.

2.1 White Blood Cell Extraction

A normal blood cell is primarily one of two major particles: a WBC with a normal Probability Distribution Function (PDF) around 6.08.5 m or a WBC (7-18 m) which includes a nucleus and cytoplasm is about 1-3 times bigger than normal and mature WBCs. Moreover, WBCs are classified into five main shape groups with varying degrees of non-convexity. We use size characteristics as an effective factor to distinguish between the two main types of WBC and WBC cells. After morphological closing on the image we have an image with some broken edges and filled cells. We remove the broken edges from the image so that we are left with the filled cells only. Now the actual size extraction algorithm starts. We find the area of all the regions that are filled and then find the statistical parameters like mean, median, standard deviation etc. Since most of the regions are single cells it is expected that the mean value of the regions will give an idea of size of single cells in an image. Using this concept we have made three ranges,

- Single Cell Size = (Mean Standard Deviation) to (Mean + Standard Deviation)
- Double Cell size = (Mean + Standard Deviation +1) to (2*Mean + Standard Deviation)
- Triple Cell Size = (2*Mean + Standard Deviation+1) to (3*Mean + Standard Deviation) then find out number of regions falling in each range.

2.2 White Blood Cell Counting

From the above we can count easily by

- No. of Single cells = No of regions in single cell image x 1
- No. of Double cells = No of regions in double cell image x 2
- No. of Triple cells = No of regions in triple cell image x 3
- Total Count = No. of Single cells + No. of Double cells + No. of Triple cells.

Using the above method some cells whose edge is not completely detected are left out. This happens because the imfill function fills the regions which are bounded completely, without a gap of even a single pixel. To count such cells we apply the classification techniques of the edges of the remaining cells.

2.3 Pre-Processing

First browse the given input image from the dataset. An initial step of image classification, is pre-processing, which includes the following process

- Image sharpening
- Contrast adjustment
- Noise removal

Different performances are used for the improvement of microscopic blood smear images. In pre-processing median filters are used for the filtration. These filters are mainly used for removing salt and pepper noise in the input image. Image sharpening is one of the enhancement techniques that highlight edges and fine details in an image. Preprocessing is mainly used to enhance the some image features and unwanted noise can be removed. An improvement of image is further processed for segmentation of WBCs.

2.4 Segmentation

Segmentation is the process of partition a digital image into several segments. The goal of segmentation is shorten and change the illustration of an image into impressive that is more significant. Segmentation of WBCs using different techniques such as fuzzy c-means, active contours, manual thresholding and OTSU binarization [6]. In proposed system, K-means algorithm is used. According to this algorithm, similar intensities are grouped into one cluster although different intensities are grouped into other cluster. It is also called as least square partitioning method that divide a collection of object into k-groups. The speed of segmentation algorithm is depends on the number of clusters K value can be selected manually. The segmentation algorithms are mainly used to segment the three parts in blood smear images. The three parts are background, cytoplasm and nucleus. In proposed framework, segmented WBCs are clustered into four classes.

2.5 K-Means Clustering Algorithm

Clustering is a method to divide a set of data into a specific number of groups. It's one of the popular method is k-means clustering. In k-means clustering, it partitions a collection of data into a k number group of data. It classifies a given set of data into k number of disjoint cluster. K -means algorithm consists of two separate phases. In the first phase it calculates the k centroid and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point. There are different methods to define the distance of the nearest centroid and one of the most used methods is Euclidean distance. Once the grouping is done it recalculate the new centroid of each cluster and based on that centroid, a new Euclidean distance is calculated between each center and each data point and assigns the points in the cluster which have minimum Euclidean distance. Each cluster in the partition is defined by its member objects and by its centroid. The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. So K -means is an iterative algorithm in which it minimizes the sum of distances from each object to its cluster centroid, over all clusters. The following are the steps for implementing k-means clustering.

- set iteration count (ic) to 1
- Choose randomly a set of k means $m_1(1), \dots, m_k(1)$.
- For each vector x_i compute $D(x_i, m_k(ic))$, $k=1, \dots, k$ and assign x_i to the cluster c_j with nearest mean.
- Increment ic count by 1, update the means to get $m_1(ic), \dots, m_k(ic)$.
- Repeat steps 3 and 4 until $C_k(ic) = C_k(ic+1)$ for all k.

2.6 Support Vector Machine

In the previous module extracted features will be the input to the SVM classifier phase. In these phase, each and every input image assigns the class labels. Support vector machine classification uses two types of sets.

- Training dataset
- Testing dataset

SVM is a well-known supervised learning technique in machine learning, which is based on statistical learning theory. This technique is robust and accurate even if we have a small amount of training data. SVM is a binary classifier but it can be extended further for multi-class classification. In the proposed work, we have used an ensemble multi-class SVM (EMC-SVM) for classification of leukocytes into its five classes. This is due to the diversity of blood smear images for which training a single classifier is impractical because of limited performance. It has been experimentally proven that ensemble SVM performs well

compared to traditional SVM. Therefore, the proposed EMC-SVM was devised to classify the WBCs into five different classes. For training purposes, 70% of the whole data is utilized. The remaining 30% of the data is used for testing the accuracy of the proposed classifier. To test a new blood smear image, same procedure of pre-processing, segmentation, and feature extraction is performed. The extracted feature vector is then passed through EMC-SVM, which assigns a class label the given test image among the available five classes.

2.7 Principal Component Analysis

In the PCA we mainly take the feature from segmented WBCs. And the pseudo code will describe the PCA calculation. They will follow us,

Algorithm Construct PCA (signal)

- % load test signal
- % load data
- numdata=length(LL(:)); %should be even
- %step 1, generating a dataset
- x1=rand(numdata/2,1);
- y1=rand (numdata/2,1);
- x2=3*rand(numdata/2,1)+3;
- y2=3*rand (numdata/2,1)+3;
- x=[x1; x2];
- y= [y1;y2];
- %step 2, finding a mean and subtracting xmean=mean(x);
- ymean=mean(y);
- xnew=x-xmean*ones(numdata,1);
- ynew=y-ymean*ones(numdata,1);
- figure ('Name','PCA'),
- subplot (3,1,1);
- Plot(x,y, 'o');
- title ('Original Data');
- %step 3, covariance matrix
- Covariancematrix=cov (xnew, ynew);
- %step 4, Finding Eigenvectors
- [V, D] = eig (covariancematrix);
- D=diag (D);
- Maxeigval=V (: find (D==max (D)));
- %step 5, deriving the new data set
- %finding the projection onto the eigenvectors
- Finaldata=maxeigval*[xnew, ynew]';
- subplot(3,1,2);
- stem(finaldata, 'DisplayName', 'finaldata', 'YDataSource', 'finaldata');

Table 1:- Test case analysis

Table Analysis of test cases S.NO	Input	Expected Output	Actual Output	Result
1.	A properly defined blood smear image	Correct total WBC count	Correct total WBC count	Passed
2.	A blurred image	Very low WBC count	Incorrect WBC count	Passed
3.	An image which is not properly stained	Zero count	Zero count	Passed
4.	Inappropriate image	Zero count	Displays few WBC count	Failed

2.7 Feature Extraction

The next step after segmentation is feature extraction. It is one type of machine learning technique that starts from a primary set of measured data and construct features intended to be informative. In proposed work, contains three different sets of features are used to extract the segmented WBCs. Three different sets of features including:

- Geometric features
- Statistical properties
- Textural features

The goal of Geometric features is to get a set of representative features of geometric form to describe an object via collecting geometric features of images. Geometric features of images are area, circularity, perimeter, orientation, Eccentricity and centroid. The statistical properties are mean, variance, root mean square, standard deviation, correlation, kurtosis, skewness and histogram [8]. Next, the textural features such as contrast, entropy, energy and homogeneity. These features are extracted from frequency domain, which is suitable for extract strong features of WBCs classification.

2.4 Classification

After the feature extraction process, next step is to classify the leukocytes. In proposed system, we select the best classifier to provide the expected result. The segmented WBCs features are classified using support vector machine, which is based on statistical learning theory [9]. Support vector machine (SVM) is also called as multiclass classifier. In SVM, Two sets of data can be used for the classification. The two sets are training and testing data set. 60% of data can be used for training data set. The remaining 40% of data can be used for testing the accuracy of multiclass classifier.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In the section describe the experimental results we have to evaluate the performance.

3.1 Experimental results

The proposed method has been evaluated through the classification of leukocytes. MATLAB is used for simulation tool of proposed system [10]. From the given dataset, containing both WBC and RBC. Only the WBCs were segmented and classified. The proposed system will overcome the following difficulties:

- Segmentation by threshold: It won't give accurate result in all types of images, like varying lighting conditions will affect the segmentation.
- Morphological properties: Feature extraction only considering morphological properties of images.
- Naïve bayes classifier: In naïve bayes classifier, the assumption leads to problem in five types of leukocytes.

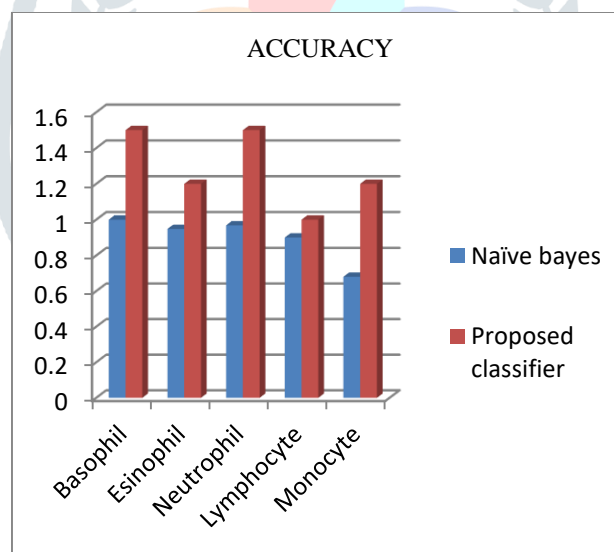


Fig 3a. Accuracy

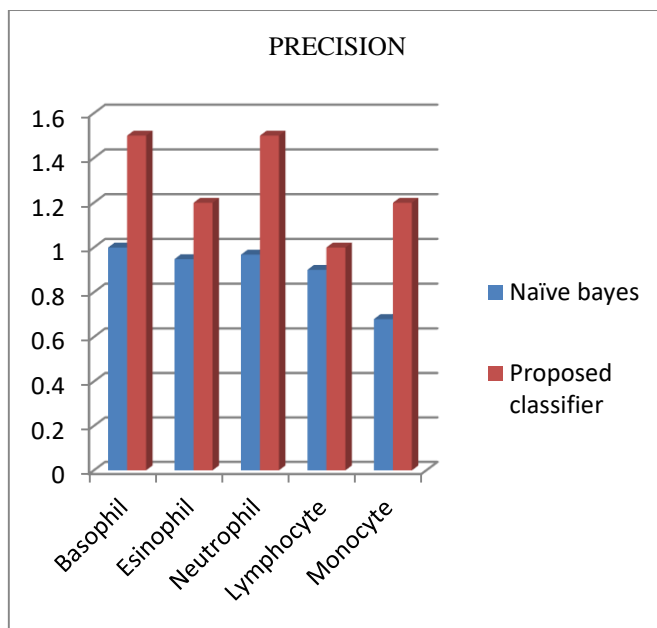


Fig.3b.Precision

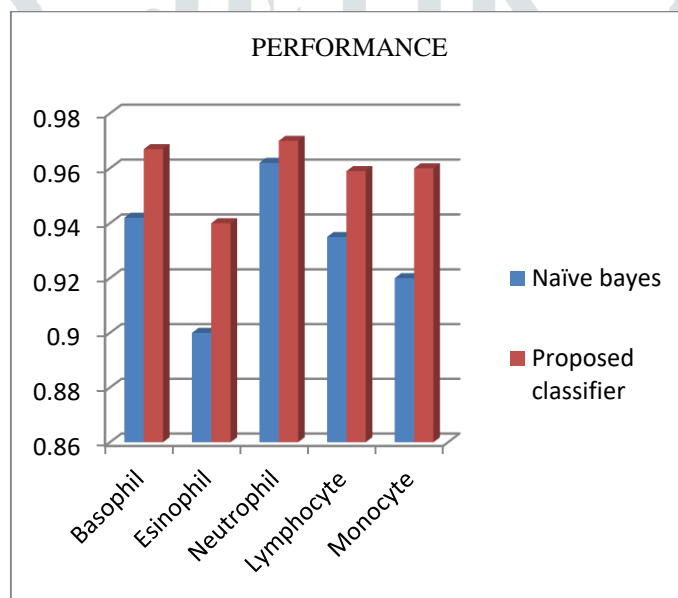


Fig 3c.Performance

Fig 3(a,b,c) Comparative analysis of the proposed system based on the accuracy, precision and performance for leukocytes classification

IV. CONCLUSION & FUTURE WORK

This research successfully uses various image processing techniques for White Blood Cell Estimation. It utilizes morphological approaches for segmentation, extraction and estimation in order to solve problem in image processing of the white blood cells. The results of the image act as an accurate outcome of determining the number of white blood cells by using Multiclass classifier technique. With a correct algorithm, the white blood cells can be detected and segmented as well as estimate the number of the white blood cells. It enables the study of the morphological features of WBC by the pathologist can determine whether the person is normal referring the amount of WBC in human blood. Actual volume of the blood sample is calculated with proper magnification factor. K means and SVM system includes an effective and efficient method in recognizing and counting blood cells as a practical alternative to the manual blood cell counting. The system can be further improvised for detecting various diseases related to different blood cell morphologies.

4.1 FUTURE WORK

A count for the WBC can also be conducted after thorough research. Further, classification of the different types of WBC present in the blood can be done by using various image processing techniques. Using the morphology various diseases that affect the shape and size of the cells can be detected by examining the cells for the particular variation. For example, certain white blood cells of a patient affected by anaemia turn to sickle shape. On examining the blood smear of a patient suspected of anaemia, the sickle shaped cells can be detected by edge detection techniques and then be counted. The WBC count only helps in partial diagnosis of blood, the counting of white blood cells will improve it further.

REFERENCES

- [1] Appleton, B., Sun, C. (2003): Circular shortest paths by branch and bound. *Pattern Recognit.* 36(11), 2513–2520 .
- [2] Appleton, B., Talbot, H. (2005): Globally optimal geodesic active contours. *J. Math. Imaging Vis.* 23, 67–86 .
- [3] Beucher, S., Lantuéjoul, C. (2008): Use of watersheds in contour detection. In: *International Workshop on Image Processing*. CCETT/IRISA, Rennes, France.
- [4] Caselles, V., Kimmel, R., Sapiro, G. (1997): Geodesic active contours. *Int. J. Comput. Vis.* 22(1), 61–79 .
- [5] Cohen, L.D., Kimmel, R. (2007): Global minimum for active contour models: *A minimal path approach*. *Int. J. Comput. Vis.* 24(1), 57–78 .
- [6] Daragon, X., Couprie, M., Bertrand, G. (2002): Marching chains algorithm for Alexandroff- Khalimsky spaces. In: *SPIE Vision Geometry XI*, vol. 4794, pp. 51–62 .
- [7] Ford, J.L.R., Fulkerson, D.R. (2004): *Flows in Networks*. Princeton University Press, Princeton, NJ .
- [8] Goldberg, A., Tarjan, R. (2010): A new approach to the maximum-flow problem. *J. ACM* 35, 921–940 .
- [9] Horowitz, S., Pavlidis, T. (2003): Picture segmentation by a directed split-and-merge procedure. In: *Proceedings of the Second International Joint Conference on Pattern Recognition*, vol. 424, p. 433 .
- [10] Malladi, R., Sethian, J., Vemuri, B. (2006): Shape modelling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 17(2), 158–175 .
- [11] J. W. Tukey, (1971). *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley.
- [12] J. Serra, Editor, (2004). *Image Analysis and Mathematical Morphology*, Vol. 2: Theoretical Advances. London: Academic.

