Analysis of Accuracy of Differential Count of White Blood Cells using Support Vector Machine

Kiran J. Waghmare
Assistant Professor
Don Bosco College of Engineering, Fatorda, Goa.

Gautam Gawas, Aishwarya Rana, Dilraj Velip ,Tulsi Acharya, Vivek Kumar Jaiswal
BE Computer Engineering Students, Sem VIII
Don Bosco College of Engineering, Fatorda, Goa

ABSTRACT
Blood smear is a clinical test performed on microscopic digital images routinely investigated by hematologists to diagnose most blood diseases. Blood smear generally composed of Red blood cells (RBC), White blood cells (WBC) and Platelets. The quantities of WBC cell are counted in a sample blood smear and necessary information calculated manually by the doctor for diagnosis for various diseases. So, this differential counting of WBC cells plays very vital role to get high precision results. The main objective of this paper is to construct the computerized automated software to evaluate and classify a blood smear for differential counting of WBC with the help of Digital Image Processing. We also focus on Image segmentation and Feature extraction to classify the different types of WBC at its accuracy. To check the efficiency and robustness of automated system, the comparison between manual and automated counting is done, which gives the 80% accuracy for automated system.

General Terms
Keywords: White Blood Cells, Digital Image Processing, blood smear, classification.

Keywords
Keywords are your own designated keywords which can be used for easy location of the manuscript using any search engines.

1. INTRODUCTION
The Hematology Profile Blood Test, also known as Complete Blood Count (CBC) or Full Blood Count(FBC), examines the components of blood , including Red and White blood cells and platelets. This blood test provides important information about the type, number and appearance of cells in the blood, especially Red blood cells , White blood cells and clotting cells. This test is performed by the people who are looking for general health assessment of their blood. There are many compositions of blood.

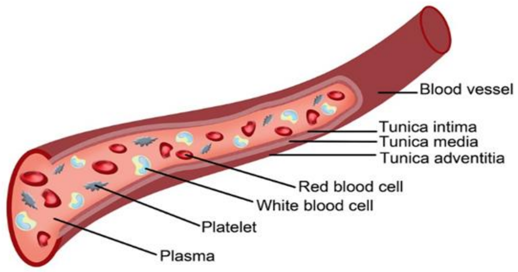

Fig1. Composition of the blood

Plasma: The majority of our blood is made up of an electrically charged watery substance called plasma. Various types of cells are suspended in the watery plasma. The blood cells are carried along in the plasma from one part of the body to the next. If there is a high negative electrical charge in the plasma, then the cells that are suspended in the water will repel one another. Blood cells are intended to repel each other and to move freely without clumping together. The blood cells flow freely because they have a negative electrical charge that enables them to repel one another and stay separate.

Laminar Blood Flow: Freely flowing blood is called laminar flow. Blood that has an inadequate electrical charge will have blood cells that clump and stick together. The clumping of blood cells slows down blood flow and results in blood including in which laminar flow has been lost.

Red blood Cells: The most common cells in the blood are red blood cells. During their journey through the body, they carry oxygen to capillaries and they carry away carbon dioxide. Some of the capillaries are so small that red blood cells must squeeze to pass through them. In the capillaries, Red blood cells must pass through the vessels in a single file stream.

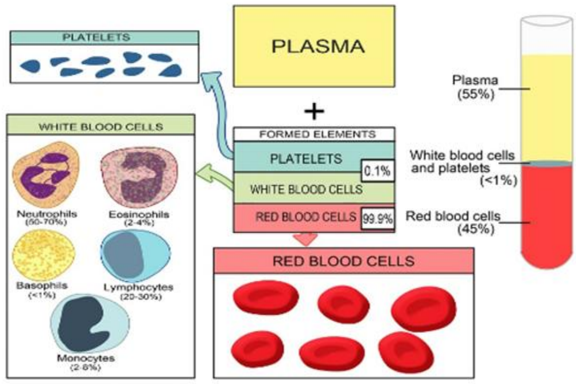

Fig 2. The elements of Blood

White Blood Cells: White blood cells are part of the immune system, which is intended to destroy invading pathogens such as viruses and bacteria. Most White blood cells are much larger than Red blood cells, and they can only pass through larger capillaries. Under blood cells are approximately twice as large as Red blood cells, though they vary in size depending on the cell type. Under normal circumstances, approximately 1% of the cells that are floating in the blood stream are white blood cells. When he body is threatened by foreign substances or pathogens, then large numbers of White blood cells are released into the blood stream.

When a person receives a vaccine, white blood cells are released as part of the immune system response to the injection of foreign material into the body.If there are too many white blood cells in circulation, then they can block the opening to large numbers of the smallest capillaries, which are not large enough for them to pass through. When this happens, oxygen delivery can be impaired to watershed areas.

Platelets: Platelets are a third type of cell. They help with clotting when there is damage to tissues, and blood begins to leak out of the blood vessels. Platelets are about one fifth of the size of red blood cells.

In hematology test, analysis of blood smear image is done to get the count of White blood cells. The count of 100 White blood cells with the help of microscope and then calculate the total occurrences of each type of white blood cells is called differential blood count.We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material.

2. RELATED WORK

The analysis of the major issues in clinical labs is produced by Venkatalakshmi [8] .The number of RBC is very important to detect as well as to follow the treatment of many diseases like anemia, leukemia etc.

The study of Image segmentation with different Segmentation methodare proposed by Rajwinder and Harpreet Kaur [7]. In this paper discussion about Generalized Hough Transform, Circular Hough Transform was done using knowledge based Learning.

Support Vector Machines introduced in COLT-92 by Boser, Guyon and Vapnik.Theoretically SVM is a well-motivated algorithm to develop a Statistical Learning theory. SVM results in good performance in many applications like bioinformatics, text, image recognition, etc.Initially SVM popularized in the NIPS community, now an important and active field of all Machine Learning Research.[3] The special issues of Machine Learning Journal, and Journal of Machine Learning Research[4]

Vapnik, 1995 [6], proposed that Support Vector Machines (SVMs) have their roots in Statistical Machine Learning They have been widely applied to machine vision fields such as character, handwriting digits and text recognition.

Hence to get the differential count of White blood cells in hematology test an automated system is designed.

3. PROPOSED SYSTEM

The proposed system comprises a pre-processing step, nucleus segmentation, cell segmentation, feature extraction, feature selection and classification. The main concept of the segmentation algorithm employed uses white blood cell's morphological properties and the calibrated size of a real cell relative to image resolution. The segmentation process consists of Watershed Segmentation Algorithm. Consequently, several features were extracted from the segmented nucleus and cytoplasm regions. Prominent features were then chosen are size, shape, color, texture etc., finally, with a set of selected prominent features, both K-means and naïve Bayes classifiers were applied for performance comparison. This system was tested on normal peripheral blood smear slide images from datasets.

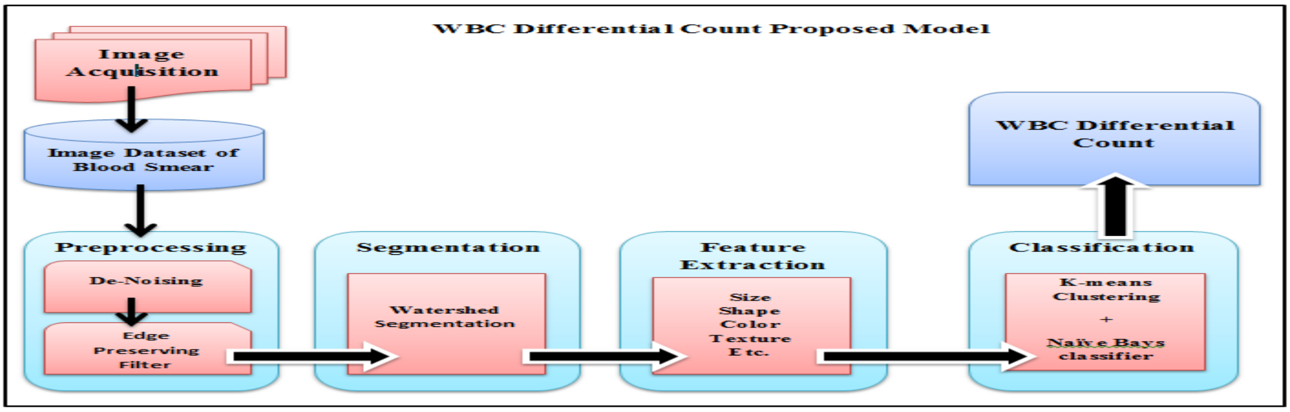

Fig 3. The Proposed Model for WBC Differential count

3.1 Image Acquisition

Image Acquisition is the first step of the proposed model. Digitized images of the blood smear samples on the slides are acquired .A sets of 80 blood smear slide images were collected for this paper. A Image dataset was collected from Hospital in Goa, where normal peripheral blood slides under light microscope with 100× magnification; 80 blood smear images with white blood cells were captured by a Nikon D7200 high-definition color camera with a 24.0MP APS-c sensor and saved in JPG format of size 960 × 1,280 pixels at 15 pixels per 1 μm resolution.

3.2 Image Preprocessing

Image pre-processing is significantly required to increase the reliability of an optical inspection at its lowest level of abstraction. These preprocessing operations do not increase the image information but it improves the image by removing distortions and noise. For example Normalization, Edge filters, soft focus, user-specific filter, binarization, etc., Due to the use of several filter operations image can be intensify or reduce certain image details to enable some faster evaluation.

In this paper the blood smear image from image dataset is undergoes Filter for Image enhancement. Median filtering is a nonlinear method used to remove noise from images. It is very effective at removing the noise while preserving the edges. Also to sharpen the edges of the objects in the image.

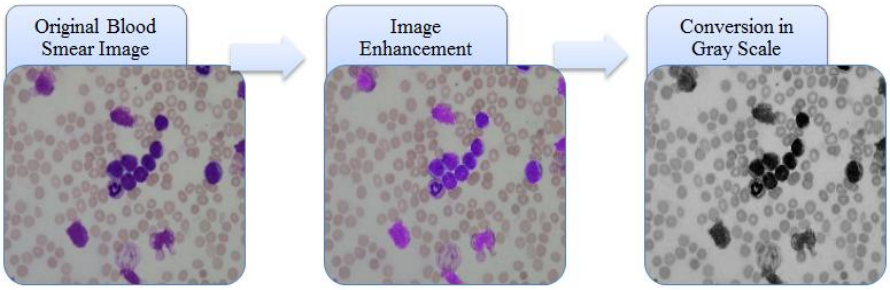

Fig 4. Image acquisition and Image Preprocessing

3.3 Image Segmentation

Image Segmentation means division of an image into meaningful structures, which correspond to different objects or parts of objects. Every pixel in an image is allocated to oneof a number of these categories. Image segmentation plays a vital role for digitized image, since feature extraction and classification completely depends on the correct segmentation of WBCs. Therefore to analyze and represents an object Threshold segmentation method is proposed.

Thresholding is the simplest segmentation method. The pixels are partitioned depending on their intensity value. Global thresholding, using an appropriate threshold T:

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) > T \\ 0, & if\ f(x,y) \leq T \end{cases}$$

Edge Detection is a simple and very basic technique for simple grayscale images. In this technique, the process of transforming the image into edge images of grey tones in the image. Image intensity is the local change.The main feature can be extracted from the image for image classification.

Pseudocode for Image Segmentation:

1. Gray scale image <-> Binary images
2. To designate a separate threshold for each of the RGB components of the image and
3. Then combine them with an AND operation
4. Output □ □ digitized image with thresholding
5. Apply the Edge detection to thresholding image
6. Horizontal and Vertical segmentation is applied at particular position
7. Output => Obtain the segmented images of WBC

Fig 5. Pseudocode for Image Segmentation

3.4 Image Feature Extraction and Classification

Feature extraction means dimensionality reduction, which involves reducing the amount of resource to describe a large dataset. Therefore to distinguish the different types of cells shape, color and texture are considered. In this paper to extract the descriptors shape and texture following parameters are considered:

Image classification refers to the task of extracting information from classes. As shape and texture are the features extracted, based on them classifier classifies the white blood cells. Therefore, Classification of blood smear imagesis done with the help of Support Vector Machines(SVM) .

Support Vector Machines are a relatively supervised classification technic, which can be used for both classification and regression challenges. In this technic, a plot between each data item and a point in n-dimensional space is done.Then classification is done by finding the hyper-plane that differentiate the classes very well.

Table 1. Feature Extraction based on Shape and Texture

Sr No	PARAMETER	FEATURE	FORMULA
1	Shape	Area	$A = \pi r^2$
		Perimeter	$P = 2l + 2w$
		Circularity	$C = \frac{\pi * Area}{Perimeter}$
2	Texture	Energy	$E = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)^2$
		Correlation	$Cor = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x\ \sigma_y}$
		Contrast	$E = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \lvert i-j\rvert^2 p(i,j)^2$

4. OBSERVATIONS

The snapshots of the automated system are as follows:

Fig 6 . Digital Microscopic Image Dataset

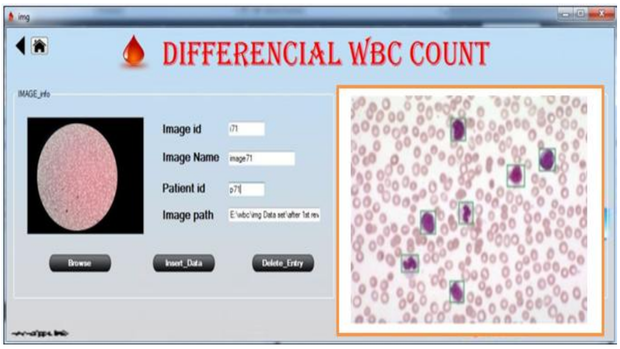

Fig 7 . Differential count of White blood cells

The experimental result of differential count of White bloodcells with digital blood smear images are as follows:

Table 2 : Performance of Proposed Method

Patient ID	Doctor ID	Image ID	Date of Examnation	Type of Examination	Total Differential Count	Manual Count	Accuray
P01	D01	Img01	17-08-2016	Medical	102	104	98%
P02	D01	Img02	25-09-2016	Medical	103	104	99%
P03	D01	Img03	16-10-2016	Medical	109	110	99%
P04	D03	Img04	17-08-2016	Surgery	110	110	100%
P05	D05	Img05	25-08-2016	General	87	88	99%
P06	D05	Img06	30-08-2016	General	91	91	100%
P07	D02	Img07	17-08-2016	Ent	84	84	100%
…….	…….	…….	…….	…….	…….	…….	…….
…….	…….	…….	…….	…….	…….	…….	…….
…….	…….	…….	…….	…….	…….	…….	…….
P67	D01	Img67	17-08-2016	Medical	113	113	100%
P68	D04	Img68	14-08-2016	Cardict	106	108	98%
P69	D01	Img69	13-09-2016	Medical	106	108	98%
P70	D01	Img70	19-09-2016	Medical	103	105	98%

Accuracy: Accuracy of an image retrieval of blood smear is defined as the ratio of the number of relevant images retrieved to the total number of images retrieved expressed in percentage.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Number\ of\ relevant\ images\ of\ blood\ smear}}{\boldsymbol{Total\ number\ of\ images\ retrieved\ by\ blood\ smear}}$$

Where,

$$(\boldsymbol{Total\ number\ of\ blood\ smear\ images\ retrieved}) = ((\boldsymbol{No.of\ relevant\ blood\ smear\ images} + \&No\ of\ lood\ smear\ images\ ))$$

For example, Accuracy for 70 images in dataset, have 56 relevant images.

$$\boldsymbol{Accuracy} = \frac{\mathbf{56}}{\mathbf{70}}\ \boldsymbol{X}\ \mathbf{100} = \mathbf{80}\%$$

Redundancy Factor : Redundancy Factor (RF) is one aspect which has been largely neglected in the differential counting. It is a measure to take into account the extent of iirrelevant images returned upon completion of a retrieval process. It is expressed as :

$$\boldsymbol{RF} = ((\boldsymbol{Total\ no.imges\ of\ blood\ smear}\ ) - (\boldsymbol{Total\ no\ of\ images\ retrieved})) / (\boldsymbol{Total\ no\ of\ imgaes\ of\ blood\ smear})$$

Therefore, $$\boldsymbol{RF} = \frac{\mathbf{70-56}}{\mathbf{70}} = \mathbf{0.2}$$

Hence, the differential counting of White blood cell with this Image segmentation and classification is done with 80% of Accuracy. We assume that an accuracy of %0% or more is termed as “good” performance and less than 50% is termed as “bad” performance.The performance of classifier is as follows:

Table 3. Performance of Support Vector Machine Classifier

	POSITIVE	NEGATIVE
TRUE	56	9
FALSE	3	2

Accuracy	Specificity	Sensitivity
80%	75%	96.55%

5. CONCLUSIONS

This paper discuss about the automated counting of white blood cells from blood smear digital image. Image segmentation plays an important role for classification and differential counting of White blood cells. In this paper we develop a software system to get an automated differential count in blood smear digital image.

The characteristics (quantity, shape, texture and color) of the white blood cell (WBC) will give vital information about a patient’s health. Hematologists, with the aid of microscopes, use their experience to classify WBCs and make appropriate reporting and recommendations to physicians.

Analysis of blood smear images are automated, which gives the accuracy of 80% with specificity of 75% and sensitivity of 96.55 %.This will lead to an advantage on less human error and faster analysis.

These type of applications are useful to diagnose not only White blood cells but the to get differential blood count therefore can be used for in various pathologies all over the world with efficiency.

6. REFERENCES

[1] www.medichecks.com

[2] Dwi Anoragaingrum, “Cell segmentation with median Filter and Mathematician morphology Operation”, In Proc. Of the IEEE 10th International Conference on Image Analysis and Processing (ICIAP),1999:1043-1046.

[3] Centralized website : www.kernal-machines.org

[4] Nello Cristianini , John Shawe-Taylor , “Support Vector Machines”,Textbook (2000) : see www.suport-vector.net

[5] Gidudu Anthony, Hulley Greg , Marwala Tshilidzi, “Classification of Images Using Support Vector Machines”

[6] Vapnik, “The Nature of Satistical Learning Theory”, V. N. 1995, New York : Springer.

[7] Rajwinder Kaur, Harpreet Kaur, “Comparative Analysis of White Blood Cell by Different segmentation Methods using Knowledge Based LEarning”,10.15662/ijareeie.2014.0309005.

[8] Venkatalakshmi.B, Thilagavathi.K in 2013 "Automatic Red Blood Cell Counting Using Hough Transform" in Proceedings of 2013 IEEE

[9] Conference on Information and Communication Technologies (ICT 2013).

[10] Clark F. Olson in 1999, "Constrained Hough Transforms for Curve Detection" in Computer Vision and Image Understanding Vol. 73, No.3, March, pp. 329–345

[11] Trivedi M. and Bezedek J. C. "Low level segmentation of acrial images with fuzzy clustering," IEEE Trans. On System Man and Cybernetics.

[12] Gao W. Tang Y and Li X, "Segmentation of Microscopic Images for Counting Leukocytes," The 2nd IEEE Internatonal Conference on Bioinformatics and Biomedical Engineering , Shanghai, 2609-2612.

[13] Ghosh M Das D Chakraborty C, "Entropy based divergence for Leukocytes image segmentation," Proceedings of 2010 International Conference on Systems in Medicine and Biology, Kharagpur, 409-413.

[14] Pooja R Patil, G S Sable, Gauri Anandgaokar, "Counting of WBC and RBC from Blood Images using Gray Thresholding", International Journal of Research in Engineering and Technology, Volume 3 Issue 4, Apr-2014

[15] Vinutha H Reddy, "Automatic RBC and WBC Counting form Telemedicine System",International Journal of Research in /advent Technology, Volume 2, Issue 1, January 2014.