

Blood count analysis using machine learning

Shraddha kshirsagar

*Department of Electronics and Communication Engineering
SVERI's COE, Pandharpur, Maharashtra, India*

Dr. Anup vibhute

*Department of Electronics and Communication Engineering
SVERI's COE, Pandharpur, Maharashtra, India*

Abstract- Blood analysis is one of the major important factor in human health as many diseases can be diagnosed based on cell count. The traditional approach of the blood cell count is very time consuming. So in this paper, we have used a machine learning approach to classify & then we have done counting of blood cells. For this purpose, we have obtained the microscopic images of blood cells from the laboratory. We have successfully classified all the blood cells using SVM machine learning image category classifier with 100% accurate results. After that by applying certain image processing techniques we have done counting of blood cells. We have also developed GUI to make the system simple & reliable. Hence in this approach we have correctly classified the image category & further got the results within less time than the traditional blood cells counting approaches.

Keywords – **Blood count, Machine learning, RBC&WBC, Classification**

I. INTRODUCTION

The complete blood count (CBC) is the blood test used to evaluate the health of people & to detect the disorders like anemia, infection and leukemia [14]. Generally, blood cell can be categorized in 4 types: Red Blood Cells (RBCs), White Blood Cells (WBCs), Platelets & Plasma [2]. WBCs are important part of immune system & are also called as Leukocyte. These cells protect the body by removing viruses & bacteria in a body. These are composed of nuclei and cytoplasm & are divided into five groups [3]. Different automated methods are developed for segmentation & classification of WBCs [10]. Features like textural, statistical, geometrical & morphological are used for the classification [4]. Some fuzzy divergence & thresholding techniques are also applied for the recognition [13].

RBCs are also known as erythrocytes. The function of RBC is to hold oxygen & collects CO₂ from the lungs to the cells of the body. Segmentation of RBCs from other blood constituents can be done by histogram thresholding & different morphological operations [8]. Artificial Neural Network (ANN) works properly for identification of clustered RBCs when sufficient amount of samples were supplied [5].

For the better diagnosis of patient health we have developed a fast & accurate system. In the past research many systems have used publically available dataset for their results but we have used real blood smear images obtained from the laboratory to get the accurate results. Machine learning approaches have the better ability to estimate, predicate & classify the blood cells than the existing systems. SVM is one of the powerful method of classification as it is having good generalization ability & extremely powerful learning procedure [4]. We have used SVM machine learning algorithm for image category classification on training & testing set images & have obtained 100% classification accuracy which describes the performance accuracy of the system. In the counting of the blood cells we have used different image processing techniques for the segmentation of images [6]. We have also used circular half transform to detect the overlapped cells. To make the system more user friendly we have also created the GUI output which will make system more reliable than existing systems. The results obtained are then compared with the laboratory results to check the performance accuracy of result. Hence proposed system works better, solve image category classification problem & gives accurate results than the past methods.

The rest of the paper is formalized as: In section II, we have discussed the past related work done in the blood cell analysis. In section III, our proposed methodology for classification & counting of blood cells is given. After that section IV, explains the performance analysis of the proposed system. The section V, describes the accuracy of system & in section VI, conclusion & future work is explained.

II. RELATED WORK

In recent years the interest of researchers in the field of medical imaging especially in blood image processing is increasing the algorithms for automated analysis of medical images. Researchers have made lot of improvements in their research by the application of Artificial intelligence, image processing & pattern recognition, & computer vision techniques to blood cell analysis [11]. There are several automated systems are developed for medical diagnosis that are helpful for the doctors to detect diseases particularly in WBC & RBC [14].

Yanhui Guo, and co-authors has developed a deep learning tool is used for image category classification. Paper represent three different approaches for WBCs identification system: DeCA approach, fine-tuned network approach, and low depth DL architecture which is called as WBCsNet. According to this paper, the proposed method can be applied for detection of leukemia disease. The author Abdullah Elen & their co-authors has discussed & compared several machine learning algorithms. Their performance is tested by creating feature vectors on training & testing set images. Razali Tomaria, & other co-authors Computer Aided System for Red Blood Cell Classification in Blood Smear Image. According to this paper the ANN frameworks can effectively identified the clustered RBC cell when sufficient sample cells were supplied. Author Vincenzo Piuri, & others presented a methodology to achieve a fully automated detection and classification of leucocytes by microscope color images. This proposed system firstly individuates in the blood image the leucocytes from the others blood cells & then it extracts morphological indexes and finally it classifies the leucocytes by a neural classifier.

III. PROPOSED METHODOLOGY FOR CLASSIFICATION & COUNTING OF BLOOD CELLS

The proposed methodology for classification & counting of blood cells can be explained in 3 steps as follows:

1. Image acquisition & database creation for image classification.
2. Feature extraction & creation of image category classifier.
3. Counting of blood cells (WBC & RBC).

Details of each step can be explained below.

1. Image Acquisition & Database creation for image classification

1.1 Image Acquisition

In the proposed method we have firstly obtained the blood smear images from the lab. The images obtained are microscopic images taken from digital microscope. The images are in .jpg format. We have obtained separated images of RBC & for the WBC identification we have used purple stained blood smear images in .jpg format.

1.2 Database creation

The obtained .jpg images from the lab are used for database creation. These database images are then further divided as training & testing images. To check the accuracy results obtained from this method are compared with the laboratory results. Following figures shows the database images collected.



Figure 5. (a) Database images for RBC counting
(b) Database images for WBC counting

2. Feature extraction & creation of image category classifier

2.1 Feature Extraction

After creation of database we have divided images as training & test set image. Then we have created 'Bag of Visual Words' to extract SURF features of an image as shown in figure 3.

2.1.1 Bag of visual words

We have created the bag of visual words which defines features of given input images. The 'bagOfFeatures' function defines the features, or visual words, by using the k-means clustering on feature descriptors of the training set images algorithm extracted from training images. In this bag of visual words histogram of visual words is created which describes given input image. Bag of features technique is depends on detection without localization. Training image classifier with bag of visual words. By extracting the feature descriptors from input images a visual vocabulary, or bag of features can be extracted. According to this we have created a bag of features as shown in figure 2. This bag of feature is formed is a cluster the of k groups which is compact & has simple characteristics. These clusters indirectly represents the given input image. Hence for the classification of image category, we have extracted 'Speeded Up Robust Features' (SURF) as it provides the greater scale variance.

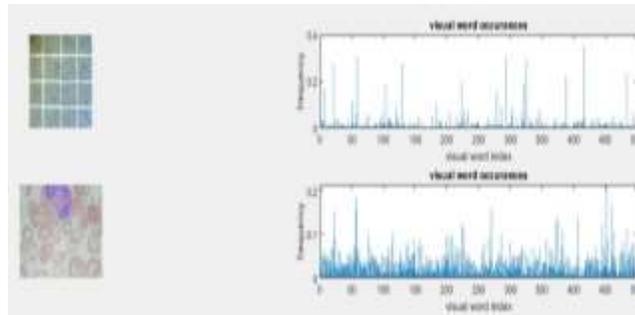


Figure 2. Bag of visual words created for input images

2.2 Creating category classifier for classification

'imageCategoryClassifier' is the function used to create category classifier. Using this function we have trained Support Vector Machine (SVM) classifier to classify given input image category. This SVM classifier uses Error Correcting Output Codes (ECOC) framework for classification. SVM is supervised machine learning algorithm which uses binary classification & regression. It creates optimal hyperplane margin to separate two classes. Following figure shows the functions used & creations of image Figure 3 indicates creation of category classifier for two image category classification. This category classifier will classify the image in WBC or RBC category.

create category classifier for classification

```
categoryClassifier = trainImageCategoryClassifier(training_set_bag);
confMat_r16 = evaluateImageCategoryClassifier(test_set);
meanID(mf(confMat_r16));

Training an image category classifier for 2 categories.
-----
* Category 1: Wbc
* Category 2: Rbc

* Sampling Features for 12 images...done.

* Finished training the category classifier. We evaluate to test the
  classifier on a test set.

Evaluating image category classifier for 2 categories.
-----
* Category 1: Wbc
* Category 2: Rbc

* Evaluating 77 images...done.
```

Figure 3. Creation of image category classifier for input images

3. Counting of blood cells (WBC & RBC)

After classification of image category we will go for counting of blood cells. The counting of blood cells includes further steps as image preprocessing, image enhancement & image post-processing.

3.1 Counting of RBC

3.1.1 Image preprocessing

Image preprocessing is the step carried out to enhance the quality of image to make it suitable to carry out further processing. In our work we have used simple image processing steps to achieve the better results. Firstly we have converted RGB image into grayscale image for processing. Figure 4 shows the RGB & converted grayscale image.

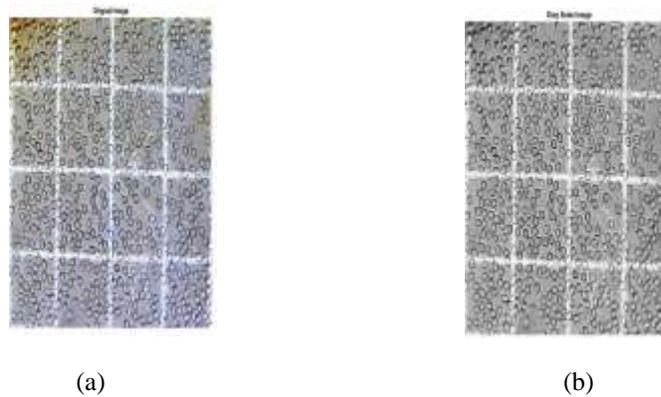


Figure 4. (a) Input image (b) Gray image

3.1.2 Image Segmentation

Image segmentation is done to extract object from background. In the image segmentation we have done thresholding to create the binary image from the grayscale image. For the edge detection we have used 'canny edge detector'. After this we done some mathematical morphological operation & cleared the borders of an image. Canny edge detector is most widely used operator as it provides good localization & improved signal-to-noise ratio. We have used mathematical morphology type of dilation. In this dilation process 'strel' type of structuring element. Generally application of this structuring element make image more brighter & hence bright regions will get increase in size & dark regions will get shrinked. 'imclearborder' function in MATLAB is used to clear the borders of image. Figure 6 will illustrate the whole process done on the grayscale image.

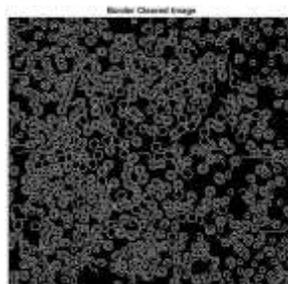


Figure 5. Border cleared image

3.1.3 Image post-processing

In the image post processing we have used 'imfill' MATLAB function for holes filling. After holes filling we have filtered the circle shaped RBCs on the basis of area of cell by using 'propfilt' MATLAB function. By taking the mask of smallest & largest circle present in an image we have filtered out an image between the defined mask area. For the extraction of mask we have defined the mask for smallest & largest circle. In this way we will get the extracted circular cells from the image. Further for counting the number of circular cells we have used concept of connected components. The connected components are the group of pixel which are having same pixel values & are connected to each other by either 4 or 8 pixel connectivity. Following figure 6 shows the extracted circular cells in an image.



Figure 6. Extracted circular RBC cells

3.2 Counting of WBC

For the extraction & counting of WBC we have used purple stained blood smear images which we have obtained from the laboratory.

3.2.1 Image preprocessing for WBC extraction

In the extraction of WBC we have firstly extracted the blue plane of from the image. We can also extract red & green plane from the RGB image. But the shape of WBC that we need for further processing is obtained better in blue plane. So that here we have used only blue plane for the extraction of WBC cells. The following figure 8 will show the given input blood smear image & extracted blue plane image.



(a)

(b)

Figure 7. (a) Input image
(b) Extracted blue plane image

3.2.2 Extraction & counting of WBC

We have used 'bwareaopen' MATLAB function to remove noise & small objects. To achieve measurement properties from image we have used 'regionprops' function. Here we have specified regionprops as 'all' which will computes all the shape measurements. We have used same concept of connected components for extracting WBC cells. Here we have used 4 pixel connectivity as we used in RBC counting. 'NumObjects' function will give us the count of WBC cells present in the image.



Figure 8. Extracted WBC cell

IV. PERFORMANCE ANALYSIS OF PROPOSED SYSTEM

Image classification performance analysis

For the checking of performance of image classification first we have created bag of features & then applied the classification algorithm on test set images. The diagonal 1s created in the confusion matrix below indicates the accuracy of classification. Diagonal 1s indicates that image classification is done correctly. Following figure10 explains the classification results of test set images.



Figure 9. Confusion matrix

V. PERFORMANCE ACCURACY FOR BLOOD CELLS COUNTING

Graphical user interface (GUI)

For making the counting of blood cells easy to handle we have also created the graphical user interface using MATLAB designer app. In this GUI we have created 2steps. In the first step we will be taking input image & then preprocess the image & in second step we will segment the image, remove the noise & small object to calculate the number of blood cells in an image. The Edit Field inserted while designing will automatically shows the number of blood cells present in that particular image. Following figures will show the GUI outputs.



(a)

(b)

Figure 10. (a) GUI output for RBC counting (b) GUI output for WBC counting

VI. ACCURACY OF THE APPROACH

We have collected the laboratory reports of RBC & WBC counting of 5 patients & compared it with our results. From this we have calculated the accuracy of our approach by using following equation.

$$\text{Accuracy} = \frac{\text{Cell count from our approach}}{\text{Cell count from laboratory}} * 100$$

Following table 1 shows the results obtained from laboratory & results obtained from our approach.

Patient no.	Results of proposed approach		Results of laboratory	
	RBC (M/ μ L)	WBC (Cells/ μ L)	RBC (M/ μ L)	WBC (cells/ μ L)
1	4.2	5200	4.2	5200
2	4.1	7200	4.1	7300
3	4.0	6200	4.2	6100
4	4.5	1700	4.5	1600
5	4.7	5000	4.7	4500

Table 1. Result Analysis

VII. REFERENCES

- [1] Siti Norul Huda, Sheikh Yazan M. Alomari, *Automatic Detection and Quantification of WBCs and RBCs*, Hindawi Publishing Corporation, 2014.
- [2] Maziz.othman, Thrbit,S.Mohammed, Alaa Ali, "Neural Network Classification of WBCs using Microscopic images", *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 5, 2017 .
- [3] Abdullah Elen, M.kamil Turan, "Classifying White Blood Cells Using Machine Learning Algorithms ", *International Journal of Engineering Research and Development* ',Issue1, volume1.
- [4] Stanislaw Osowski, Robert Siroi'c, Tomasz Markiewicz, and Krzysztof Siwek, "Application of Support Vector Machine and Genetic Algorithm for Improved Blood Cell Recognition" IEEE Transaction on Instrumentation and Measurement,VOL.58, 2009
- [5] Razali Tomaria, Wan Nurshazwani Wan Zakariaa, Muhammad Mahadi Abdul Jamila," Computer Aided System for Red Blood Cell Classification in Blood Smear Image", International Conference on Robot PRIDE 2013-2014 - Medical and Rehabilitation Robotics and Instrumentation, ConfPRIDE 2013-2014.
- [6] Akshaya P. Sahastrabuddhe1, Dr. Sayyad D. Ajj "Blood group Detection and RBC, WBC Counting: An Image Processing Approach", *International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 5 Issue 10 Oct. 2016*.
- [7] Yanhui Guo , K.M. Amin A.I. Shahin a, "White blood cells identification system based on convolutional Deep Neural Network Learning," *Computer Methods and Programs in Biomedicine,Elsevier*, Nov. 2017.
- [8] Venkatalakshmi.B, Thilagavathi.K, "Automatic Red Blood Cell Counting Using Hough Transform" IEEE Transaction on Information and Communication Technologies 2013.
- [9] M.Alagappan, B.Banu Rekha, R.Arun, "Extreme Learning Machine (ELM) based Automated Identification and Classification of white blood cells", *International Journal on Machine Learning and Cybernetics, Springer-Verlag, Vol.2, pp. 107-122, 2011*.
- [10] Herbert Ramoser, Vincent Laurain, Horst Bischof, " Leukocyte segmentation and classification in blood-smear images", IEEE Transaction Engineering in Medicine and Biology 27th Annual Conference 2005.
- [11] Muhammad Imran Razzak, SaeedaNaz, "Microscopic Blood Smear Segmentation and Classification using Deep Contour Aware CNN and Extreme Machine Learning", IEEE Transaction on Computer Vision and Pattern Recognition, 2017.
- [12] Vincenzo Piuri, Fabio Scotti, "Morphological Classification of Blood Leucocytes by Microscope Images", IEEE International Conference on Computational Intelligence far Measurement Systems and Applications Boston, USA, 16 July 2004.
- [13] Devkumar Dasa, Chandan Chakrabortya, Ajoy K. Rayb,Madhumala Ghosh, *Automated leukocyte recognition using fuzzy divergence*, Science direct, Ed.: Elsevier Ltd, 2010.
- [14] Wael Hassan Gomaa Fahad Kamal Alsheref, "Blood Diseases Detection using Classical Machine Learning Algorithms," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. Vol. 10, 2019