Digital Imaging of data using Machine Learning

[1]Mrs. Renuka V Jadhav, [2]Mr. Vinayak Shashikant Jadhav

[1]Assistant Professor, [2]Head of Department
[1]Bharati Vidyapeeth College of Engineering Kolhapur
[2]Sahyadri Polytechnic Sawarde

***Abstract*- Using the mathematical approaches and the algorithm that are required for their recognition of images. The uses of jai application in image processing delay the process of image processing. With this new approach of algorithm a new invention in the identification of the data, now and more clear digits data set appears into sight can be appeared. But the problem lies in the behavior and capturing of those images of data sets base, we design a more sophisticated algorithm for digitally representation model based on machine learning, collection of raw data images. This paper consist of different approaches to an off methods using different machine learning techniques. Several machine learning algorithms like Linear Regression Algorithm,Logistic Regression Algorithm,Decision Tree,SVM,Naïve Bayes, ,KNN Multilayer Perception, Support Vector Machines, Convolutional Neural Networks, and many more. The main purpose is detection and quantification of images to recognize the effective and efficient method for recognition of the pattern. The paper shows different classification algorithms have their different accuracy.**

Keywords: **Machine Learning KNN RBC WBC Histogram**

1. INTRODUCTION

The microscopy analysis of images is extremely important in both the medical and the computer science fields. Many research problems are related to the analysis of microscopy images, such as complete blood count (CBC) tests [1] and the analysis of blood smears, which is considered the first step in detecting and diagnosing malaria, leukemia, and anemia. Additionally, during a complete physical exam a series of tests are performed. One of these tests is the CBC, which is used to evaluate the composition and concentration of all cellular blood components. The CBC determines red blood cell (RBC) counts, white blood cell (WBC) counts, platelet counts, hemoglobin (HB) measurements, and mean red blood cell volumes [2]. CBC tests and the analysis of blood smear images help to evaluate, diagnose, and monitor various health conditions, such as anemia, leukemia, infections, and allergic conditions [3]. For blood disorders, such as anemia, which is based on HB level, the production and destruction of red blood cells are evaluated. In red blood cell disorder such as anemia, other red cell indices such as (mean cell volume) MCV, mean cell hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), RBC, and red blood cell distribution width (RDW or RCDW) are evaluated to narrow down on the causes of anemia. If the red cell indices are suggested of iron deficiency anemia (IDA), further tests to confirm the IDA will be done. In normal blood, red blood cell (RBC) counts range from 4.2 to 5.9 million cells per square centimeter. High RBC counts can be indicative of serious medical conditions, such as heart, lung, or kidney disease. Primary or secondary polycythemia in polycythemia HB is also raised; a bone marrow disorder also causes high RBC counts [2]. Normal WBC counts range from 4,500 to 10,000 WBCs per microliter of blood [4]. High WBC counts (above 30000 cells per microliter) indicate an infection, systemic illness, inflammation, allergy, leukemia, or burn induced tissue injury. If leukemia is suspected, analysis of blood smear is done to look for morphology of the leukemic cells and followed by bone marrow examinations [2–4]. For platelets, which are small blood cell fragments that assist in blood clotting, normal counts range from 150,000 to 450,000 platelets per microliter. In patients with low platelet count such as in patients with dengue infection, their platelet count is monitored closely and the value is within critical level, the patient might need platelet transfusion [2]. Generally, any abnormal blood smear reading indicates an infection or disease. Malaria and Babesia are parasites that infect RBCs; the analysis of thin blood smear remains the gold standard for diagnosis in such disease [5]. Microscopy images are also still used for early diagnosis, analysis, and count of some blood disorder such as sickle cell anemia and leukemia, before confirming it with other laboratory tests. However, manual or visual quantification of parasitemia in thin blood films and WBCs in leukemia is an extremely tedious, subjective, and time-consuming task with a high probability of counting error [6, 7]. An accurate segmentation and counting mechanism that gathers information about the distribution of microscopic particles may help diagnose abnormalities during clinical analysis. Our objective in this paper is to develop and validate an algorithm that segments and automatically counts red and white blood cells in microscopy images. The ground truth of the images was determined by experts. For the evaluation, quantitative analyses were performed on the segmentation results based on the ground truth, and the *F*-measure method was used to confirm accuracy. In the following sections of this paper, we will summarize related work on the segmentation and counting of RBCs and WBCs, present the methodology used (Section 3), discuss the results and experiments , and review

2. RELATED WORK

Many researchers have investigated blood cell segmentation and counting. Some researchers [5, 8–10] used morphological operations and thresholding to do the segmentation and counting. Berge et al. [5] presented an approach based on a morphological method and iterative threshold techniques. Segmentation was performed on red blood cells, which included clumped cells, and boundary curvatures were used to construct a Delaunay triangulation. They used real microscopy images prepared in the laboratory, and the ground truth was determined by a laboratory expert. A 2.8% difference was calculated between the manual and automatic counting of red blood cells. Their method tolerated a degree of overlapping, but in cases with a high degree of overlapping cells,

the cells were unable to be detected. Additionally, the iterative threshold method was unable to detect faint red cells. Damahe et al. [8] used the S and V image components of a HSV color model with Zack's thresholding technique for cell segmentation. Thresholding combined with a sequential edge-linking algorithm was used to increase segmentation accuracy. The experimental dataset and blood cell images were collected from the Dhruv Pathology Lab and the CDC-DPDx, respectively, and the ground truth was determined by experts. The dataset size was limited. Several RBCs that were detected possessed holes. Additional preprocessing steps should have been implemented to increase the accuracy. Panchbhai and Vishal [9] proposed an automated analysis that counted red blood cells and detected malaria parasites in thin blood smear samples. The green color layer was processed to count all of the RBCs; segmentation was performed on the infected RBCs using Otsu thresholding. A histogram was used to determine the optimal threshold. CDC datasets were used for the experimental part, and the ground truth was determined by a pathology expert who compared their results with the manual counts. However, because the detection algorithm used morphology and thresholding, their method was unable to detect clumped and overlapping cells. Khan et al. [10] proposed a method to count WBCs, RBCs, and platelets. Several preprocessing steps were performed before converting the image to binary. Segmentation and cell counting were performed based on the optimal threshold value, which was determined from a histogram. They achieved 95% accuracy with their proposed method compared to manual counting and a hematology analyzer. However, this method is unable to detect overlapping cells. When using iterative thresholds, the probability of losing useful information from the image is high; this decreases the accuracy of segmentation. Nguyen et al. [11] used distance transform to solve the overlapping cells problem; they proposed a method that concentrated on clumped cells. First, they assigned central points based on a distance transform. The optimal center points were selected by checking the degree of boundary covering the center point, and the average size of a cell was estimated by the extraction of a single cell. Then an algorithm was developed that used a single cell mask to split the cells. Their dataset ground truth was labeled by experts, and the accuracy of the proposed method using F-measurements was 93.5% and 82%. Clumped cells were tolerated, but the cells had to be regularly shaped and focused at a high magnification. Not all blood cells have regular cell shapes, and this is especially true if the blood cells are diseased. Therefore, cell detection methods should be able to detect irregular cells. Additionally, because of noise the performance of their method was not good. Rhodes and Bai [12] presented a circle detection method using specific properties of Gabor wavelet filter to detect the image features such as circularity. It is able to extract radius wraps around the origin and the plane wave radiates from the center of the filter. They test their proposed method on synthetic images and real microscopic images, which allow a certain degree of overlapping cells. The results were 91.3% and 87% of the cells detected in the two microscopic test images. Since blood cells are approximately circular shape, circle detection algorithms can still handle the challenge of blood cells detection. Hough Transform is considered as one of the most known algorithms for line and circle detection. It was developed by Richard Duda and Peter Hart in 1972 [13]; they called it as a generalized Hough Transform after the related 1962 patent of Paul Hough [13]. Hough Transforms maps every edge pixel into parameter space and use conventional HT to detect lines, circles, or any other parametric shape. In this paper, we concentrate on circle detection algorithms. There are many techniques used for circle detection. Many HT-based algorithms that detect circles were developed using different methods. Yip et al. [14] reduced the accumulator array to enhance computation time and improve memory consumption. Other methods use pixel gradient information [15, 16] to reduce computation time and the accumulator array. Ho and Chen used the geometric properties of circles to improve performance [17]. Xu et al. [18, 19] developed the randomized Hough Transform (RHT), which randomly selects three noncollinear edge pixels (a, b, r), maps them into parameter space, and requires less computation time and memory storage compared with Standard HT methods. A simple voting strategy in the accumulator is used to collect evidence and determine the existence of a circle. Chung and Huang [13] presented a method that can substitute various shape detection algorithms. This method enhanced the speed of the original algorithm. They applied their method in RHT and randomized circle detection (RCD) algorithms to detect lines, circles, and ellipses. Their method presented good results when compared with original methods. Chiu et al. [20] presented a fast randomized Hough Transform method for detecting circles, to improve RHT which is less efficient in complex images due to its probability usage problem. They pick one random edge pixel from the image, and it is considered as a seed point. Then, a checking method was developed to observe if this selected seed point is lying on a true circle or otherwise. The checking criterion is based on finding two other points whose distances are the same from the selected seed point by using a $W \times W$ window centered by the selected seed point. This method enhanced the probability to find relevant points on a true circle. They have proven that using one random selected point's probability is sufficient in comparison to three random selected points [13–19]. Chen and Chung [21] presented a circular algorithm called RCD. They claimed that RCD outperformed other most efficient Hough Transform based algorithms [13–19]. Regardless of accumulator usage as in [13–19], RCD works by randomly selecting four edge pixels from the whole image. Then, these pixels are examined if they are noncollinear and it will proceed to form a candidate circle. RCD determines that circle is a possible circle based on distance criteria. After finding its center and radius, it checks the number of pixels lying on the boundary of this possible circle. This checking criterion is performed by calculating the distance between all edge pixels in the image and the boundary of this possible circle. If this distance is lesser than a fixed threshold value, then it will be considered as a boundary of this possible circle. Finally, another fixed threshold value has also been used to decide a true circle or otherwise based on number of edge pixels lying on a possible circle's boundary. Other related issues on RCD are less efficient when dealing with huge image size consisting of a high number of edge pixels. RCD also requires four selected edge pixels randomly from the whole image, which causes low probability forming a true circle. Furthermore, RCD has a drawback in terms of its fixed number of iterations in which it is highly correlated to the image texture. In addition, RCD acquires many parameters and threshold values to be predefined and it ignores irregular circles. Since the Hough Transform presented a good performance in different fields, many researchers used Hough transform method for detecting circle when performing RBC's and WBC's calculation task in the microscopic images. Mahmood et applied Hough transform method for counting the RBC's and WBC's for the microscopic images. In the first step, they converted the source image to $L*a*b$ color space model and performed color segmentation process of red and white cells based on feature lightness over the L channel ranging between 130–150 and 80–100, respectively. Then, the second step begins with applying some morphological operators which are

followed by applying Canny operator for edge detection process. Lastly, they performed cell detection and counting using Circular Hough Transform. The dataset was composed of 108 images from the "Acute Lymphoblastic Leukemia Image Database for Image Processing" database that was established and maintained by Labati et . Depending on the processing and type of cells being analyzed, they achieved an accuracy ranging from 64% to 87%. The Hough Transform consumed memory storage and required a long computation time to determine a large range of accepted radii. Additionally, the ability to tolerate a high degree of overlapping and irregular cells was limited; therefore, accuracy was not high. Mahmood and Mansor examined 10 image samples of normal blood cells; image transformed to the HSV color space, and then Saturation or "S" channel was selected to proceed with image analysis. Morphological operators and thresholding method were used over S channel for cell segmentation. They used Circular Hough Transform to investigate the circularity feature of the red blood cells in order to perform detection and counting. Their proposed method achieved approximately 96% of accuracy rate in comparison to manual counting. Extracting and counting normal cells are simple tasks if the detected cells are normal cells and consist of small number of overlapping cells with regular shape. For this kind of easy case, the idea of using Hough Transform can be very helpful because it can produce a good performance. and shape features. Those Venkatalakshmi and Thilagavathi have also applied circular Hough Transform method to count the RBCs from microscopic images after performing preprocessing steps such as HSV transformation, S channel extraction, histogram thresholding morphological operations, XOR logical operation, and Canny edge detection. However, again, this proposed idea is less tolerant to any overlapping cells or irregular cells' shape. Later, Maitra et al. presented a composition method to extract red cell from five microscopic images; these steps include spatial smoothing and filtering, adaptive histogram equalization, and edge detection. Similar to, they used basic Circular Hough Transform to detect the red cells based on prior information such as size and shape features. Those methods employ classic Hough Transform method to detect the blood cells which inherits some drawbacks, for example, required more computation, high memory consumption, and less ability in detecting overlapping cells or irregular cells. Finally, Cuevas et al. presented a method which is not Hough Transform based to detect white blood cells. They used a combination of circle detection with electromagnetism-like optimization from the edge map image. This method tolerated noise. In blood smear images, the number of white cells was small compared with the number of red cells; therefore, small degree of overlapping white cells can be detected with their method. However, they did not test the method on clumping cells. Their results were compared with other methods, and their method demonstrated good accuracy.

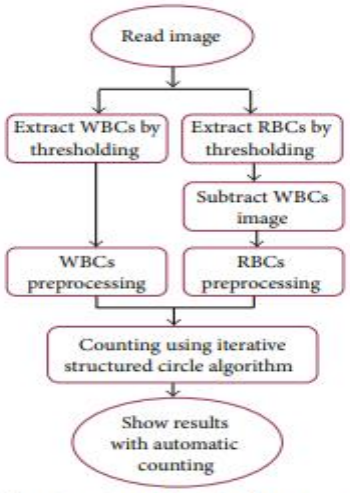

FIGURE 1: General methodology for the proposed method.

3. METHODOLOGY

The proposed method was developed to analyze microscopic images of blood smears by segmenting and counting both WBCs and RBCs. The segmentation is based on thresholding and morphological operations, and then counting is based on the circularity feature of the blood cells extracted using an iterative structured circle detection algorithm. A new technique for binary images based on the fundamentals of RCD has been proposed and used for counting RBCs and WBCs. Therefore, the original image is separated into two images; the first image contains RBCs only and the second image contains WBCs; this step has been done using thresholding. We study the histogram for 20 sample gray scale images, and we find out the best thresholding values to extract WBCs and RBCs; values were 64 and 140, respectively. After cells separation, each image is preprocessed using morphology operators to obtain the edge image using Canny operator. Then, an iterative structured circle detection algorithm is used to count cells in each image. Figure 1shows the proposed method for WBC and RBC segmentation. Preprocessing. The proposed method for cell segmentation works with edge images. Microscopy images of blood smears are colored images, and several steps are required to prepare the image before extraction of the edge image. In our proposed method, the cells were separated by type and distinct preprocessing steps were developed for WBCs and RBCs separately. Preprocessing for WBCs. At this stage, white blood cells are extracted as a separate image, and the red blood cells have varying intensities; therefore, it is preferable to develop separate preprocessing steps for each cell type. Figure 2 shows the overall preprocessing steps for WBCs. To remove RBCs from the image, the RGB image was converted into grayscale image by eliminating the Hue and saturation information while retaining its luminance, as shown in Figure , and then converted the image to binary using thresholding using a threshold value ; visualize the WBCs, as shown in Figure 3(c). Some undesired holes appeared in the cells, and the morphology operator, fill holes were used to remove them. The complementary images before the holes were filled, as shown in Figure 3(d). This image eroded to reduce the number of

overlapping cells, as shown in Figure 4(a). Because the boundary of the cells is required, the holes were filled to improve the edge detection. Figure 4(b) shows the image after the holes filled. The Canny operator was used to visualize the cell edges, as shown in Figure 4(c). After edge detection, some undesired pixels appeared that affect the segmentation process. These pixels were removed using an open morphology operator, as shown in Figure 4(d). Preprocessing for RBCs. In this stage, preprocessing was performed on the red blood cells after removing the white blood cells from the image. Figure 5 shows the overall preprocessing steps for the RBCs. First, the original image is converted into grayscale image by eliminating the Hue and saturation information while retaining its luminance.

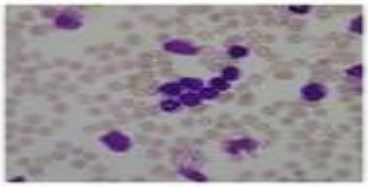

Figure 3: (a) Original blood image, (b) gray image, (c) binary image using thresholding, and (d) complement image.This image eroded to reduce the number of overlapping cells, as shown in Figure 4(a). Because the boundary of the cells is required, the holes were filled to improve the edge detection. Figure 4(b) shows the image after the holes filled. The Canny operator was used to visualize the cell edges, as shown in Figure 4(c). After edge detection, some undesired pixels appeared that affect the segmentation process. These pixels were removed using an open morphology operator, as shown in Figure 4(d).

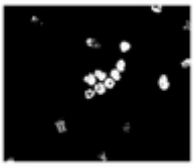

Figure 4: (a) Eroded image, (b) image after filling holes, (c) edge detection using a Canny at high magnification, and (d) image after removing the noise at high magnification.

Preprocessing for RBCs

In this stage, preprocessing was performed on the red blood cells after removing the white blood cells from the image. Figure 5 shows the overall preprocessing steps for the RBCs. First, the original image is converted into grayscale image by eliminating the Hue and saturation information while retaining its luminance.

Figure 5: Preprocessing steps for RBCs.

Then, image was converted to binary using thresholding value of (140), to visualize all of the red and white blood cells in the image, as shown in Figure 6(a). To remove the white blood cells from the image, the complementary white cells image was taken and subtracted from the first image to obtain only the red cells, as shown in Figure 6(b).

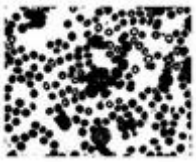

Figure 6: (a) Binary image using thresholding, (b) RBCs after removing the WBC's by subtracting the two images, and (c) image after filling holes.

When the image was converted to binary and the white blood cells were removed, undesired holes were created. These holes disturb the solidness of the object. Therefore, morphological operators are used to fill the holes, as shown in Figure 6(c).

A morphological step using an erosion operator is used to reduce the overlap between cells, as shown in Figure 7(a). Canny edge detection method is used to obtain the edge image, as shown in Figure 7(b). Some undesired pixels appeared after edge detection. These pixels represent either platelets or noise, and it will subsequently affect the segmentation process. Therefore, these pixels are removed using a morphology operator on the binary image, as shown in Figure 7(c). After these preprocessing steps, the image was prepared for input to our proposed circle detection algorithm.

The Proposed Method for Counting

The basic idea for the proposed method was derived from the RCD algorithm. RCD algorithm ignores accumulator capability that has been introduced by Hough Transform method. Several modifications were made to the basic RCD algorithm to solve the initialization problem when using big images with high number of pixels. The methods modified to detecting irregular circles,

selecting the optimal circle from the candidate circles, determining the number iterations in a fully dynamic way to enhance the algorithm detection and running time, and improving the detection of overlapping cells. The workflow of the proposed method is shown in Figure 8.

Figure 8: Proposed method workflow.

Initialization Problem. The proposed method partitions the edge image based on 8-neighbor connected components in order to overcome the initialization problem caused in basic RCD. We eventually divide the whole image into small partitions and we consider each partition as an input image before entering into our iterative structured circle detection algorithm that employs local randomization step.

Regarding to this initialization problem, as we highlighted earlier that selecting four pixels globally (from the whole image) can reduce the probability of finding true circle, time consuming, and needs of a high number of iterations to find all true circles. This can be easily illustrated as if cell A is to be detected, and the four selected pixels are chosen in Red Cross mark. Therefore, we simplify this process by introducing four pixel selections locally (within each partition image). Assume that cell A is to be detected in the partition image, and then these four local pixels are chosen randomly from the partition image as depicted. This local randomization process repeats until all partition images are visited.

CONCLUSION:

Considering various algorithm of machine learning like CNN, KNN and SVM with different data and applications histograms vectors, we found different variation among the classes considering factors accuracy and timing. Accuracy can be reformed because it relies upon on the examination of information and there may be always a risk to enhance the accuracy of those models if the dimensions of data set is will increase. each set of rules as its very own accuracy and time intake. If the energy of CPU modifications to GPU specific algorithm can carry out with higher accuracy and much less time and better result can be acquired. The performance of the classifier may be measured in phrases of potential to discover a condition nicely, the percentage of proper effects, range of positive outcomes from the technique of classification as false positives and ability to exclude.

REFERENCES:

[1] M. Habibzadeh, A. Krzyzak, and T. Fevens, “Application of pattern recognition techniques for the analysis of thin blood smears images,” Journal of Medical Informatics & Technologies, vol. 18, 2011.
[2] M. M. Osman, “Normal reference value of blood cell count, red, white and platelet of Khartoum North Area,” Al Neelain Medical Journal, vol. 3, no. 8, 2013.
[3] (NIH) National Institutes of Health, “A service of the U.S. National Library of Medicine,” http://www.nlm.nih.gov/medlineplus/ency/article/003642.htm.
[4] http://www.buzzle.com/articles/high-white-blood-cell-count .html.
[5] H. Berge, D. Taylor, S. Krishnan, and T. S. Douglas, “Improved red blood cell counting in thin blood smears,” in Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI ’11), pp. 204–207, Chicago, Ill, USA, April 2011.
[6] S. S. Savkare and S. P. Narote, “Automatic classification of normal and infected blood cells for parasitemia detection,” International Journal of Computer Science and Network Security, vol. 11, no. 2, pp. 94–97, 2011.
[7] J. M. Sharif, M. F. Miswan, M. A. Ngadi, S. Hj, and M. Mahadi, “Red blood cell segmentation using masking and watershed algorithm: a preliminary study,” in Proceedings of the International Conference on Biomedical Engineering (ICoBE ’12), pp. 258–262, Penang, Malaysia, February 2012.
[8] L. B. Damahe, P. G. Student, R. G. College, and R. G. C. Engg, “Segmentation based approach to detect parasites and RBCs in blood cell images,” International Journal of Computer Science and Applications, vol. 4, no. 2, 2011.
[9] L. B. D. Panchbhai and V. Vishal, “RBCs and parasites segmentation from thin smear blood cell images,” International Journal of Image, Graphics and Signal Processing, vol. 10, no. 10, pp. 54– 60, 2012.
[10] S. Khan, A. Khan, and A. Naseem, “An accurate and cost effective approach to blood cell count,” International Journal of Computer Applications, vol. 50, no. 1, pp. 18–24, 2012.