Applied Soft Computing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique

Pramit Ghosh [a,*], Debotosh Bhattacharjee [b], Mita Nasipuri [b]

[a] Department of Computer Science and Engineering, RCC Institute of Information Technology, Kolkata 700015, India
[b] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form
11 November 2015
Accepted 23 December 2015
Available online xxx

Keywords:
Boundary derivative
Euclidean distance
Fuzzy classification
HSI color model
Region growing
Texture

ABSTRACT

Total count and differential count of leukocytes or white blood cells (WBC) in blood samples are very important pathological factors for diagnosing a disease. There are not enough pathological infrastructures in the remote places of India and other developing countries. The objective of this work is to design a system, compatible with telemedicine, for automatic calculation of the total count and differential count of WBC from the blood smear slides. Hemocytometer based WBC counting provides more accurate result than manual counting, but hemocytometer preparation process needs expertise. As this device is targeted for remote places, blood smear technique is adopted to reduce the overhead of the operator. In the proposed system, microscopic images of blood smear sample are processed to highlight the WBC for segmentation. Region segmentation procedure involves background scaling and redundant region elimination from the region set. After segmentation, the more accurate region boundary is restored by using gradient based region growing with neighbourhood influence. Individual regions are separately classified on the basis of shape, size, color and texture features independently using different fuzzy and non-fuzzy techniques. A final decision is taken by combining these classification results, which is a kind of hybridization. A set of rules has been generated for making final classification decision based on outputs from various classifiers. The sensitivity and specificity of the system are found to be 96.4% and 79.6%, respectively on a database of 150 blood smear slides collected from different health centres of Kolkata Municipal Corporation, Kolkata, India.

1. Introduction

Human blood cells are mainly divided into three categories, namely Red blood cells (RBC) or Erythrocytes, White blood cells (WBC) or Leukocytes and Platelets or Thrombocytes. The main composition of RBC is haemoglobin, which primarily carries oxygen to living body cells and collect carbon dioxide from them. RBC have a lifetime of 120 days on an average. WBCs take care of the immune system that defends the body against both infectious diseases and foreign materials. Their life span is 3–4 days in the human body [1,2].

Platelets or Thrombocytes are tiny in size, 2–3 μm in diameter. They are irregular in shape and look like cell fragments. Platelets discharge thread-like fibbers to form clots that involved in haemostasis. The average lifetime of a platelet is normally just 5–9 days. Platelets release a multitude of growth factors which play significant roles in the repair and regeneration of connective tissues. If the number of platelets is too low, excessive bleeding may occur. However, if the number of platelets is too high, blood clots can form thrombosis that may obstruct blood vessels [3].

WBC or Leukocytes mainly protect our body against infectious diseases. They are created in the bone marrow and attack bacteria, viruses, and germs that enter into human body. There are five major types of WBC. These are Lymphocyte, Monocyte, Neutrophil, Eosinophil and Basophil. In a normal adult body, there are 4000–10,000 WBC per microliter of blood. Increase or decrease in the number of WBC in blood is an indication of an infection somewhere in the human body. The average percentage of each type of WBC in the blood are Neutrophil – 50–70%, Eosinophil – 1–4%, Basophil – 1%, Monocyte – 6% and Lymphocyte – 20–40% [4,5].

Based on the granularity of nucleus, Neutrophils, Eosinophils, and Basophils are grouped into a category, named granulocytes. These type of WBC contain digestive enzymes. The granularity of Basophils is the highest among them, Eosinophils have orange-red granules and Neutrophils have a faint blue-pink color.

* Corresponding author. Tel.: +91 9474575828.
E-mail addresses: pramitghosh2002@yahoo.co.in (P. Ghosh), debotoshb@hotmail.com (D. Bhattacharjee), mitanasipuri@gmail.com (M. Nasipuri).

http://dx.doi.org/10.1016/j.asoc.2015.12.038
1568-4946/

Neutrophils are one of the body's main defences against bacteria. They kill bacteria by ingesting them. Eosinophils also kill parasites and have a role in allergic actions in our body. Lymphocytes fight against viral and some bacterial infections by directly attacking the antibodies. Monocytes are the largest among the white blood cells. They clean our blood by eating foreign particles, bacteria, and dead Neutrophils, etc. Basophils release two chemicals, histamine, and heparin. Histamine reduces the allergic reactions, and heparin is an anticoagulant chemical, which prevents clotting of blood and helps bringing more blood to a damaged area in our body.

1.1. The necessity of WBC counting

The WBC count indicates the total number of WBC and the percentage of each type of WBC in a person's blood. This WBC counting is used to determine a variety of illnesses. For example, Leukopenia, HIV, radiation therapy, liver and spleen diseases cause very low WBC count [6]. WBC count goes very high due to Leukocytosis, anaemia, stress, asthma, etc. [7,8]. The percentage of eosinophils increases when patient suffers from allergies and parasitic infections. Bacterial and fungi infection increases the density of Neutrophils, whereas sepsis and Aplastic anemia reduces Neutrophil count. Tuberculosis and other chronic infections increase the Monocytes-RBC ratio. So depending on the WBC count doctor will recommend a treatment plan for the patient. It is very difficult to arrange sufficient health infrastructure in the remote areas of a developing country. Telemedicine partially solves the lack of physician, but pathological infrastructures are not adequate in the remote areas. Sometimes it becomes very difficult to diagnose a disease properly due to the absence of a pathological report. A microscope based automatic WBC counting system compatible with telemedicine will be very helpful in those cases [9,10].

1.2. Review of existing works

Literature survey reveals that very few research works have been done in identification of WBC in blood smear images out of them some note worthy works are discussed here. Jiang et al. proposed a WBC segmentation technique based on histogram along with scale-space filtering and watershed clustering but they did not discuss anything about the procedure for distinguishing WBC from Protozoa (like Plasmodium) [11].

Bergen et al. described a Level set based technique to identify Leucocytes in a blood smear image [12]. However, they do not focus on the classification of different types of WBC. A WBC segmentation technique is proposed by Dorini et al. where the Selfdual Multiscale Morphological Toggle (SMMT) [13] approach is used. Their technique emphasizes to find WBC cell boundary (cytoplasm boundary) more accurately. The focus of their work is to detect only the Leucocytes, not on the classification of the different types of Leucocytes.

1.3. Motivation

A number of automatic WBC counting systems exist like Automatic blood analyser [14], haematology analyser [15], etc., but their works are based on some chemical processes. They also require a considerable amount of maintenances and recurring cost. On the other hand, the techniques discussed in the review section are developed in the academic interest and do not address the real life practical issues or attempt to assist the doctors with better diagnosis. For example, the method proposed by Kin Jiang, has not considered the artefact and Protozoa; because they use perfectly processed blood Smear slides. This is not possible in the real life situations. Dorini et al. emphasizes to find WBC cell boundary (cytoplasm boundary) more accurately, which has no additional benefits in the detection of WBC.

The objective of this work is to design a complete WBC counting system, which is based on analysis of microscopic images of blood smear slides and it can be implemented using the existing manual WBC counting system with minor modification. This system is aimed to be deployed in the remote areas of developing countries as a supporting aid to telemedicine system and any person with school education would be able to operate it.

The organization of the paper is as follows. Section 2 describes the detailed design methodology of the system. Section 3 reports the performance analysis of the system and comparison with other existing systems. The next section focuses on future scope of the work and concludes the paper.

2. The design principals and methods

The work focuses on the automation of counting of different types of white blood cells in a blood sample from the digital images of the blood smear slide. Olympus CX21i [16] microscope fitted with a CCD camera is used to capture microscopic images of blood smear slides. Leishman stain is used to stain the nucleuses of WBC, platelets and other parasites or bacteria (if they are present) in the blood samples. RBCs and platelets can be easily distinguished considering their shapes and sizes. RBC is concave, and it does not have any nucleus; Platelets are tiny in size [4]. A mechanical setup is used to change the slide position under the objective lenses of the microscope. According to WBC counting guidelines, images of more than 100 fields of a single blood smear slide need to be examined for better estimation of WBC count. For this reason, the automatic mechanical slide movement setup is required. This automatic slide movement is implemented by using two stepper motors [17]. NEMA 23 type stepper motors are used for the mechanical stage movement of the microscope. TB 6560 [18] stepper motor driver controller is used to control stepper motor more accurately. TB 6560 supports the micro-steps [19] to control the rotation of the shaft of the motor precisely by 0.1125° for each input pulse, and it is independent of the pulse width.

The proposed technique considers shape, size of the nucleus, color of the cytoplasm and texture features to eliminate cells other than WBC and classify different types of WBC. Each of these features are self-sufficient to classify each type of WBC. However, in this work, independent feature based WBC classification techniques are combined to reduce the false positive and false negative rates. Hence, it is a kind of hybrid approach. Prior to the classification of WBC, the WBC regions are detected. The WBC region detection process is common to all the independent feature based classification techniques. Fig. 1 shows the block diagram of the complete system. Pre-processing is common to all the WBC classification techniques that are discussed here. Initial detection of regions, containing WBC using color features is the first step of pre-processing. After initial region segmentation, wrongly detected regions are removed. Once the actual region has been detected then, the gradient based region detection technique is applied to obtain more accurate boundary points of the nucleus. Once the nucleus is detected, the next phase is to classify the WBC type. The shape feature, and size feature are independently used to classify all the WBC types except Neutrophil and Eosinophil. In those cases color feature of cytoplasm is used to categorize them. In parallel with these classifications, texture feature strongly classifies the artefact, malaria parasite, and Basophil. However, texture feature is also capable of classifying other WBC types.

2.1. Initial region segmentation

Leishman stain colors the nucleus materials of WBC with violet in the blood smear slide. In RGB color model, violet color is

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

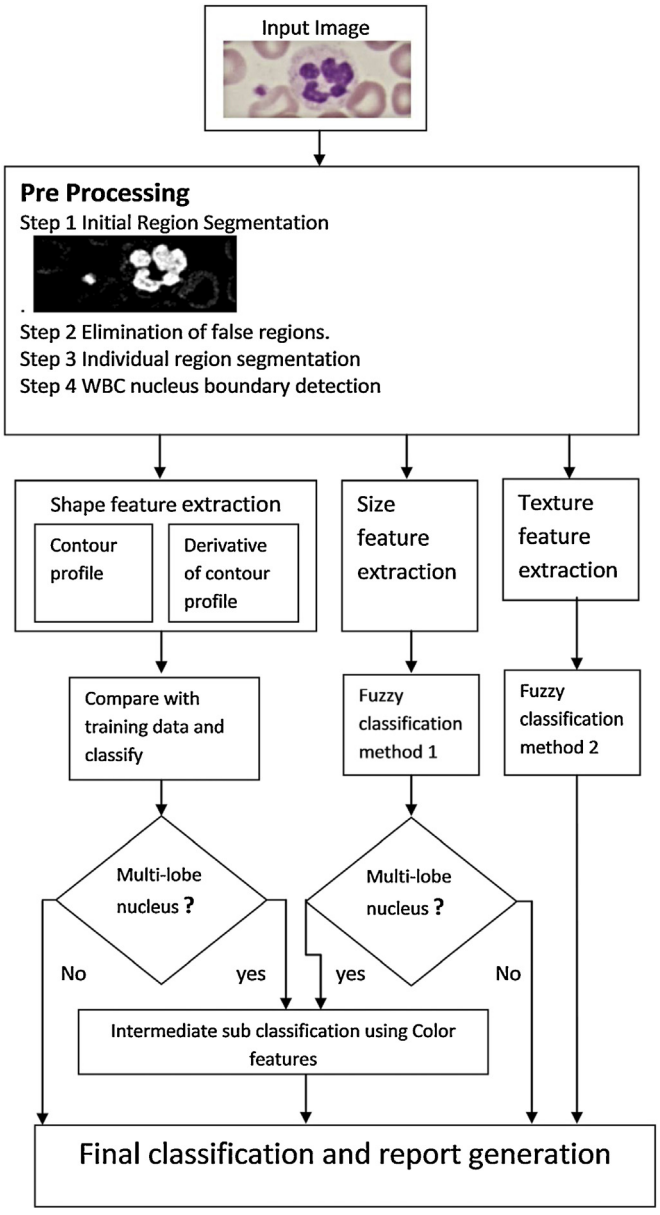

Fig. 1. Block diagram of the hybrid WBC classification system.

characterized by a very high blue component value, a very low green component value with red component lying in between. Fig. 2 shows the microscopic image of a blood smear slide. In this image, a Neutrophil with multi-lobe nucleolus along with transparent cytoplasm is present in the middle. Violet dots are platelets. One small circular object with violet color in the upper middle portion of the image is a premature WBC. The reddish concave disc-like cells are RBCs.

2.1.1. Highlighting of the WBC region

The Leishman stained nucleolus of WBCs looked violet. These violet regions need to be segmented [20] for feature extraction and classification [21]. Prior to the segmentation, these violet regions are highlighted for better segmentation [22].

It is clear from Fig. 2 is that most of the regions of the microscopic image is either white or red. Hence, white or red regions are suppressed first to highlight violet regions. White pixels have equal values of its three fundamental color components red, green and blue. Hence by examining the values of RGB components, the white and red regions are suppressed.

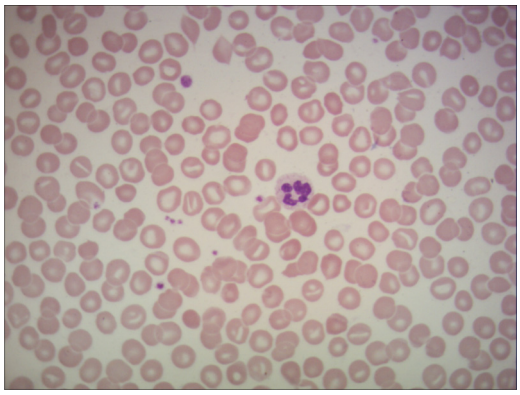

Fig. 2. Microscopic images of blood samples.

Leishman stained regions of the input image was highlighted after applying Algorithm 1. Microscopic color image of the blood smear is given as input and it returns a gray image where Leishman stained objects are highlighted as shown in Fig. 3.

Algorithm 1. To highlight Leishman stained region.

Input: The color image is represented in the form of three matrices namely red, green and blue. These three matrices contain the intensities of red, green and blue components respectively.
Output: A matrix where Leishman stained regions are highlighted.
Step
1: The difference between green matrix and blue matrix is computed and the result is stored the result in $I1$ matrix. $I1(j,k) = Green(j,k) - Blue(j,k)$ where $j = 1, 2, 3, \ldots$ number of rows and $k = 1, 2, 3, \ldots$ number of columns. The white regions are suppressed in $I1$ matrix.
2: Olympus CX21i microscope, a lens is used to focus light into sample slide. These causes non-uniform illumination in the image; the center point of the visible field has high intensity, whereas periphery has less intensity comparatively. Top-Hat transform is applied to minimize this non-uniform background in $I1$ matrix.
3: Mean intensity value mid1 is calculated upon $I1$ matrix using. Eq. (1) to determine whether scaling is required or not.

$$mid1 = \frac{1}{row * col} * \sum_{i=1,j=1}^{row,col} I1(i,j) \quad (1)$$

where $I1(i,j)$ denotes the intensity value at the (i,j)th element of the $I1$ matrix, "row" and "col" are used to indicate the total number of rows and total number of columns of $I1$ matrix. Scaling of the intensity levels is applied when mid1 value is less than the middle value, which is the average of maximum and minimum intensity values in $I1$. 4: In the scaling process the intensity values in $I1$ matrix need to be mapped in between 0 and $L-1$, where L denotes the number of intensity levels. In the scaling process, all intensity levels are mapped into a new value using Eq. (2). $P(k)$ denotes the probability for the intensity value k. $P(k)$ is the ratio between the total number of elements with intensity value k and the total number of elements in $I1$ matrix. In the scaling process, the old intensity value k will be mapped to a new intensity value $H\_new(k)$ as given in Eq. (2).

$$H\_new(k) = floor((L-1) * \sum_{n=0}^{k} P(n)) \quad here \quad k = 0, 1, 2, \ldots, L-1 \quad (2)$$

5: The difference between Red matrix and Blue matrix is stored in $I2$ matrix.
6: Test $I2$ matrix to determine whether scaling is required or not, and if required then scaling is applied on it. This process is same as in step 3 and step 4.
7: The Leishman stained regions are highlighted by subtracting $I2$ from $I1$ and the result is stored in $I3$ matrix.
8: Stop

Using the notation of set theory, the WBC regions and background pixels in the image I3 is represented as follows:

Pixel_set ={x: x is a vector representing the coordinates of a pixel. The total number of elements in pixel_set is equal to the total number of elements in I3 matrix (described in Algorithm 1).}

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

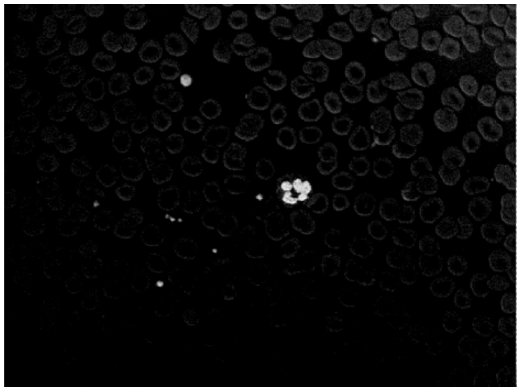

Fig. 3. Leishman stained regions are highlighted.

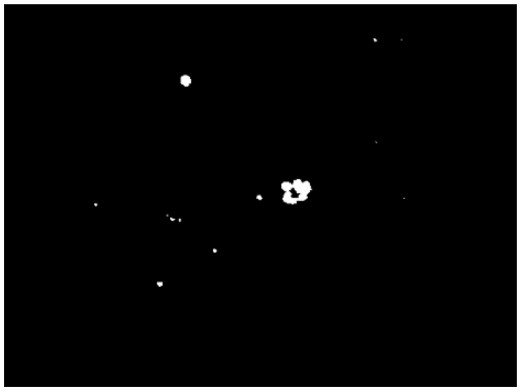

Fig. 4. Binary image of I3 using a threshold value obtained in Algorithm 2.

Region_point_set = {y: y is a vector whose elements are row number and column number of a pixel. This set contains all the pixels, which may contain pixels of WBC. }

Background_point_set = { z: z is a vector whose elements are row number and column number of a pixel. This set contains all the pixels that may not contain pixels of WBC.}

Thus, Pixel_set = Region_point_set $\bigcup$ Background_point_set.

Region_point_set = {$p \in$ Pixel_set: $p \notin$ Background_point_set }.

Region_point_set $\bigcap$ Background_point_set = ø.

The Region_point_set is the set of pixels that may contain WBC. Algorithm 2 is used to find the elements of Region_point_set by binarizing I3.

Fig. 3 shows that most of the pixels represent background. So normal Otsu's technique based binarization is not suitable to segment the WBC regions [23]. Algorithm 2 calculates a threshold value after discarding 90% of the background pixels. It calculates the threshold value iteratively and at each iteration the elements of the input matrix I3 is partitioned into two groups viz. Region_point_set and Background_point_set based on the threshold value. The algorithm stops when there is no change in the threshold value.

Algorithm 2. To binarize of the input image into foreground and background.

Input: $I3$ matrix, obtained from Algorithm 1, denoted as Pixel_set.
Output: The coordinates of the pixels of the regions that may contain WBC, denoted as. Region_point_set.
Step
1: Find out the number at pixels for each intensity level using Eq. (3); where $f(i)$ denotes the number of pixels for ith gray value. n denotes the maximum intensity value in $I3$ matrix.

$$f(i) = \frac{1}{i} * \sum_{r=1,c=1}^{row,col} I3(r,c) \quad when \quad I3(r,c) == i \ (3)$$

Where, row and col are used to denote the total number of rows and columns in $I3$ matrix. $i = 1, 2, 3, \ldots, n$
2: Discard 90% of the background pixels with low intensity using $f(i)$ values and consider the rest of the pixels of $I3$ matrix (obtained from Algorithm 1). Find the median value of the remaining 10% pixels and this value will be the initial threshold value in the next step.
3: Partition rest of the pixels into two groups using a threshold value. These two groups are denoted by P_1 and P_2. The elements of P_1 group have less intensity value than the threshold value, and P_2's elements have intensity values higher than or equal to the threshold value.
4: Calculate new threshold = (1/2)*(mean intensity value in P_1 of the elements in P_1 + $meanintensityvalueoftheelementsinP_2$). Repeat the steps 3 and 4 until there is no change between the old and new threshold values.
5: Convert the input matrix into binary using the final threshold value. The coordinates of all points whose value is greater than the final threshold value are stored as result in Region_point_set.
6: Stop

The binary output matrix obtained after applying Algorithm 2 is shown in Fig. 4.

2.2. Elimination of false regions

After applying Algorithm 2, the regions that may contain WBC are segmented. There are small dot like regions containing few number of pixels, which represent noises caused by the pigmentation of the stain. Moreover, a single WBC region may be divided into small sub-regions due to staining and other problems. This happens especially for the multi-lobe nucleus like Basophil, Eosinophil, and Neutrophil.

Morphological operator, erosion [24], is applied to remove small dots and after that dilation operator is used to merge adjacent sub-regions. This combination of erosion and dilation eliminates false regions and recombines the fragmented WBC regions. These morphological operations, however, deform the actual shape of the contours. This deformation is not accepted in shape-based classification. As a consequence, the next step focuses on segmenting each region accurately and finding its actual perimeter.

2.3. Individual region segmentation

The next step is to segment each of the regions for feature extraction and classification. Using the notations of set theory the region segmentation process can be defined as follows.

Single_Region_Set = {P: P is a subset of Region_point_Set (obtained in Algorithm 2). Elements of set P is a vector containing row and column numbers of a pixel. P contains the coordinates of all pixels of a single region. There is no common element between any two Single_Region_Sets. All elements of set P are connected with at least one other element of that set with eight connectivity}.

$P_1 \bigcup P_2 \bigcup P_3 \cdots \bigcup P_n$ = Region_point_set where $P_1, P_2, \ldots, P_n$ are all Single_Region_Sets, n is the total number of Single_Region_Sets.

$P_i \bigcap P_j = \phi$ where P_i, P_j are any two Single_Region_Set and value of i, j may be 1, 2, 3, …, n, and $i \neq j$. Each component of a Single_Region_Set contains the coordinates of a single region that may contain WBC, and they are found by applying following Algorithm 3.

Algorithm 3. To segment the foreground region of the image obtained after application of Algorithm 2 into a number of single regions in that image.

Input: The binary matrix with small false regions removed, denoted by Region_point_set.
Output: The coordinates of the pixels in each single region denoted as Single_Region_Set.
Step
1: Pick any pixel from the Region_point_set as seed point.

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

2: Find all other coordinates of the pixels in the Region_point_set those belong to the same single region containing that seed point, which is implemented using a stack and 8-connectivity property. Pixel those obtained are stored in a list. Each time a new pixel is obtained, this list is checked to avoid repeated exploration of same pixel. Continue until no such pixel co-ordinates is found.
3: The list obtained in step 2, contains the points representing a single region. This list is denoted as a Single_Region_Set.
4: Repeat step 2 to step 3 using a new seed value, which belongs to a separate region and create another Single_Region_Set. This iterative process will continue until all disjoint regions are extracted.
5: Stop

Algorithm 3 segments each disjoint region but the boundary of the regions are deformed due to erosion and dilation. The next step is to find more accurate boundary points of each of these single regions. More accurate boundary detection provides a more accurate result. Algorithm 4 is a gradient based contour region growing technique that is faster than other region growing techniques like snake algorithm [25–27].

Algorithm 4. To find accurate individual region boundary based on gradient value.

Input: $I3$ matrix and coordinates of pixels within the isolated single regions denoted by Single_Region_set.
Output: A refined Single_Region_set containing coordinates of the pixels of a region with the more accurate contour.
Step
1: Apply gradient operator, Sobel upon $I3$ and store the result in Grad_matrix and erode the Single_Region_set to shrink the boundary that will grow next.
2: Find out the pixels just outside the boundary points of an eroded Single_Region_Set $P1$.
Boundary_points_of_P1 = {M: M is a vector containing coordinates pixel of each points just outside the boundary. An element M is 8 connected with at least one element of $P1$.}
$Boundary\_points\_of\_P1 \subseteq (Region\_set)^C$
$Boundary\_points\_of\_P1 \subseteq Background\_set$
3: Sequence the boundary points of $P1$ using 8-connectivity rules and store the coordinates in ROW and COL vectors.
4: Find median gradient values from Grad_matrix, for those points which are members of $P1$ and store the result in VAL.
For each boundary coordinate, whose coordinates are stored in ROW and COL vectors, perform the following step.
Add the boundary points with the Single_region if and only if the gradient values of the boundary point (whose coordinates are supplied by ROW, COL vectors) and its adjacent boundary points are almost equal with VAL.
Remove the boundary point from Background_set and add the boundary point into Single_Region_Set.
5: Repeat the region growing process in Step 2 to Step 4, for all the Single_region_set.
6: Stop

After applying Algorithm 4, more accurate region boundary is obtained. The next step is feature extraction for each disjoint region. Fig. 5 shows the segmented regions and Fig. 6 is the corresponding output after application of Algorithm 4.

2.4. Feature extraction and object recognition

Feature extraction [28–30] is the next step to be performed for identification of the WBC regions. Proper classification of the isolated regions depends on the selection of appropriate features. The Leishman stained regions are broadly classified into three groups, namely, protozoa, WBC and platelets. Platelets are small in size. Hence, they are easily eliminated by examining the number of elements in the corresponding sub-region set. Sometimes the blood samples that are examined may be infected with the malarial parasites. The nucleus of *Plasmodium* species gets stained with Leishman stain. However, they are easily distinguished by their shape and size. In ring stage of malaria, the size of the parasite is very small, and is hosted within RBC. The size of Gametocyte stage is comparatively large, and it is almost same size with the nucleus region of the Neutrophil. The Neutrophil has no proper shape, but the

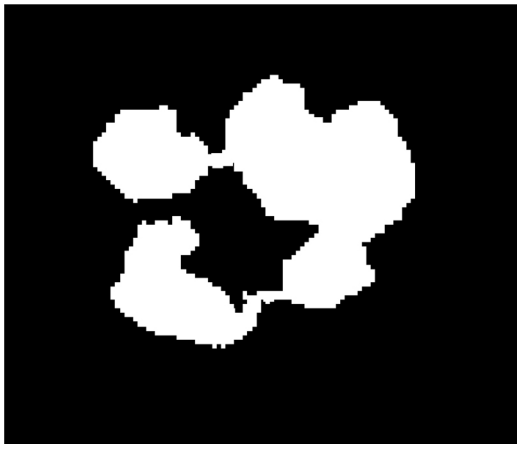

Fig. 5. The initial region that may contain WBC.

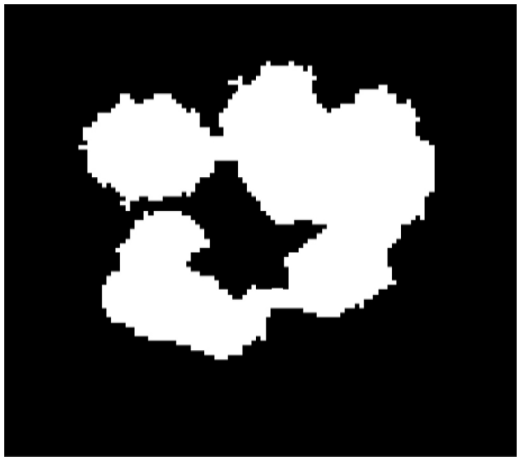

Fig. 6. More accurate region after application of Algorithm 4.

malaria parasite in gametocyte stage has a properly defined shape like circular, oval or cylindrical. Hence by examining shape and texture, parasites and Neutrophils are easily distinguished. Shape, size, color and texture features are extracted from individual region denoted as Single_region_set for classifications. These features are sufficient to classify the type of WBC, but they are combined to reduce misclassification.

2.4.1. Shape feature based classification

For extraction of shape feature, two separate independent approaches are used. One is the derivative [31] based approach, and other is contour profile based approach. In both of these cases, the signatures of the region boundary points are obtained. In the derivative approach, the boundary points of the region are extracted sequentially, and the derivative or slope of the contour of the region is calculated using Eq. (4).

$$F'(x[i], y[i]) = \frac{(y[i+1] - y[i])}{(x[i+1] - x[i])} \quad (4)$$

where, $x[i]$, $y[i]$ denotes the coordinates of ith boundary points. $F'(x[i], y[i])$ is the derivative at the ith boundary point. If the derivative is gradually increasing or decreasing then, the contour shape is the oval or circular type. However, in case of irregular shapes like the nucleus of Basophil, Eosinophil or Neutrophil, this value changes randomly. It is analyzed or compared with previously stored sample data using Eq. (5). This comparison will provide a set of feature values on which the decision will be taken.

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

Let us consider two derivative sets Z and $T1$. Z denotes the boundary point derivative set obtained from test contour, and $T1$ is the boundary point derivative from the training set. Derivative of all boundary points are not calculated, only k selected (typically $k = 200$) boundary points are considered. Thus, the number of elements in Z and $T1$ are same and they are denoted as $Z(i)$ and $T1(i)$ respectively where $i = 1, 2, \ldots, k$. $R(d)$ denotes the correlated value between Z and $T1$ with a shift d, which is defined in Eq. (5) as follows.

$$R(d) = \frac{\sum_i^k [(Z(i) - mean(Z)) \times (T1(i-d) - mean(T1))]}{\sqrt{\sum_i^k (Z(i) - mean(Z))^2} \times \sqrt{\sum_i^k (T1(i-d) - mean(T1))^2}} \quad (5)$$

$R(d)$ is calculated for all $d = 0, 1, 2, \ldots, k-1$. which results in a cross correlation series. The Z and $T1$ are considered as circular to avoid out of range index like $i - d \leq 0$. When the index is zero or negative, then it is replaced by $k + index$.

The contour profile based method calculates the center of the mass of the region using Eqs. (6) and (7).

$$Row\_centre = \frac{1}{n}\sum_{i=1}^{n} row(i) \quad (6)$$

$$Col\_centre = \frac{1}{n}\sum_{i=1}^{n} col(i) \quad (7)$$

n is the number of pixels in the region being examined. row and col are the vectors, containing the coordinates of the pixels within the region. Then the Euclidean distance between each boundary point and the center is calculated. Distances for all the boundary points of a region are used to form a contour profile of that region. Length of the contour profile vector varies in size. So this feature can distinguish between the Gametocytes stage of *Plasmodium vivax* and Lymphocyte because *Plasmodium vivax* is small in size but their shape is circular. This signature of the contour is analyzed with respect to previously stored data. Algorithm 5 is used to explain the shape feature extraction process.

Algorithm 5. To compute shape feature for a particular region from its boundary points and to classify the region on the basis of shape feature.

Input: A Single_region_set.
Output: Features that measure the similarity between different predetermined training samples of different classes and classification of the region.
Step
1: Consider an isolated single region denoted by Single_region_set $P1$.
2: The region is said to be the region of platelets or ring stage of the malarial parasite when a number of points in the region is less than a predetermined value.
3: Sequence the boundary points of the region using 8-connectivity rules and store the coordinates in a vector namely Coordinate_set.
4: Take samples from Coordinate_set at regular intervals to suppress small changes in boundary curvature. The sampling interval depends on the ratio of a total number of points in Coordinate_set and the total number of points to be selected. Store the sampled points in the Sub_coordinate_set vector. The number of elements in Sub_coordinate_set is predetermined. So that the Sub_coordinate_sets of all Single_region_sets have the same number of elements.
5: Calculate the derivative between two consecutive points from Sub_coordinate_set using Eq. (4).
6: Compare the data obtained in step 5 with training data T_1, T_2, T_3, T_4 and T_5 using Eq. (5), where T_1, T_2, T_3, T_4 and T_5 are the training data sets of the boundary derivatives for Neutrophil, Lymphocyte, Monocyte, Gametocyte stage of *Plasmodium vivax* and gametocyte stage of *Plasmodium falciparum* respectively. The corresponding results of correlations, namely RT_1, RT_2, RT_3, RT_4 and RT_5 are generated. Each of RT_1, RT_2, RT_3, RT_4 and RT_5 is a vector of length k, where k is the number of elements in Sub_coordinate_set.
7: Find the maximum values from each of the correlation vectors viz. RT_1, RT_2, RT_3, RT_4 and RT_5 and store them in F_1, F_2, F_3, F_4 and F_5 respectively. These values will be used to calculate the result.
8: Calculate the centre of the contour using Eqs. (6) and (7)
9: Calculate the Euclidean distances between the centre point and detected boundary points obtained in step 4. 10: Compare the data obtained in step 9 with training contour profile data sets UT_1, UT_2, UT_3, UT_4 and UT_5 using Eq. (5), where UT_1, UT_2, UT_3, UT_4 and UT_5 are the contour profile data sets of Neutrophil, Lymphocyte, Monocyte, Gametocyte stage of *Plasmodium vivax* and gametocytes stage of *Plasmodium falciparum* respectively. Store the output correlation vector sets into RUT_1, RUT_2, RUT_3, RUT_4 and RUT_5. Euclidean distances, obtained in step 9, are different for different contour segments. So this feature can distinguish between the Gametocyte stage of *Plasmodium vivax* and Lymphocytes because *Plasmodium vivax* is small in size but their shape is circular.
11: Find the maximum values from the correlation vectors RUT_1, RUT_2, RUT_3, RUT_4 and RUT_5 derived in step 10 and store them in F_6, F_7, F_8, F_9 and F_{10} respectively.
12: compute the scores $F_{11} = F_1 + F_6$, $F_{12} = F_2 + F_7$, $F_{13} = F_3 + F_8$, $F_{14} = F_4 + F_9$, $F_{15} = F_5 + F_{10}$
13: Based on highest score among F_{11}, F_{12}, F_{13}, F_{14} and F_{15} the classification decision is taken. In the classification F_{11}, F_{12}, F_{13}, F_{14} and F_{15} represents Neutrophil, Lymphocyte, Monocyte, Gametocyte stage of *Plasmodium vivax* and Gametocyte stage of *Plasmodium falciparum* respectively.
14: Return the decision along with the confidence value. $confidence = \frac{maximum(F_{11},F_{12},F_{13},F_{14},F_{15})}{\sum(F_{11},F_{12},F_{13},F_{14},F_{15})}$ 15: Stop

Two different shape features are considered here to strengthen the shape estimation.

2.4.2. Classification based on size feature

Platelets, malarial parasites, Lymphocyte, Monocyte and multi-lobe nucleus have different size ratio. There is no hard threshold to classify them on the basis of their sizes and size ratios [32]. Size of a region is defined as the total number of pixels in the selected region and size ratio is defined as the ratio between the size of the selected region and the size of the bounding box of that region. The fuzzy rule-based system can classify a sample region more accurately [33,34]. Rule based system is used here for initial group selection. In this technique, size of a region is examined, and if it is less than a predetermined threshold, then it is decided that the region represents either platelets or ring stage of malarial parasites. If the size is large then fuzzy based classification is used, where fuzzy membership function $f(X)$ is build using training set, and it is defined as

$$f(X) = \begin{cases} 0, & \text{if } X < \min(DSet) \text{ OR } X > \max(DSet) \\ \sum_{i=1}^{n} Yi * \prod_{i=1, j \neq i}^{n} \frac{X - Xi}{Xi - Xj}, & \text{otherwise.} \end{cases} \quad (8)$$

where X is the input sample value, $DSet$ is the training data set with n number of data. Each element of the $DSet$ contains two values (Xi, Yi) where Xi is used to denote input feature value, and Yi denotes the belongingness of Xi to the class represented by $DSet$.

Four separate fuzzy membership functions of the variable, size ratio, are built for the four classes, namely, malarial parasites, Lymphocyte, Monocyte and the multi-lobe WBC nucleus from training data set. For an unknown size ratio, these four fuzzy membership functions are applied separately, and it is assigned that class for which $f(X)$ has the maximum value. Algorithm 6 is used to find the size ratio for different WBC types and classify them.

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

Algorithm 6. To compute the size ratio of WBC region and to classify it on the basis of size ratio.

Input: A Single_Region_Set is given as input. This set contains the row and column numbers of the pixels constituting a single region.
Output: Classification of the region.
Step
1: Find the number of elements in the Single_Region_Set and store it in F_7 variable.
2: If the number, obtained in step 1, is less than a threshold value (this threshold value is the 2/3 of a single RBC size) then the region is considered as a very small region, and the region may represent platelet or ring stage of malaria. Purwar, Shah et al. proposed a method to detect Red Blood Cell (Erythrocyte) for their research work regarding malarial detection [35]. This method is used to find the size of a single RBC (Erythrocyte).
3: Find the maximum and minimum row numbers from Single_Region_Set and store them in maximum_row and minimum_row respectively.
Height = maximum_row − minimum_row
4: Find the maximum and minimum column numbers from Single_Region_Set and store them in maximum_column and minimum_column respectively.
Width = maximum_column − minimum_column
5: Bounding_Box_Area = Height * Width
6: size_ratio (F_8) = F_7 Bounding_Box_Area. Size_Ratio is quite high for Lymphocyte and Monocyte, but it is low for multi-lobe nucleus. For an unknown sample, fuzzy membership values for belonging to either of the four classes, namely, malarial parasites, lymphocyte, monocyte and multi-lobe nucleus are computed on the basis of Size_Ratio using Eq. (8) and training data sets. The training set for a particular class is a set of pairs (*Xi*, *Yi*) where, *Xi* is the Size_Ratio and *Yi* is the corresponding belongingness into that class. The final decision is taken on the basis of highest value obtained from fuzzy membership functions.
7: Return the decision with confidence value. maximum of the membership values $Confidence = \frac{\text{maximum of the membership values.}}{\text{sum of the membership values.}}$
8: Stop

2.4.3. Color feature based classification

Size and shape features are not able to distinguish between Neutrophils and Eosinophils; both have multi-lobe nucleus. The only difference between them is the color of cytoplasm [36]. Color feature based classification is only applicable when multi-lobe nucleus is detected on the basis of size and shape features. The color of the cytoplasm of Neutrophils is transparent, but it is reddish for Eosinophils. Hue from HSI [24] color model is used to classify the color of cytoplasm. The classification of Neutrophils and Eosinophils on the basis of cytoplasm color are discussed in Algorithm 7.

Algorithm 7. To Classify Neutrophils and Eosinophils on the basis of cytoplasm color.

Input: A Single_Region_Set representing a WBC region along with the input color image.
Output: Identification of the type of WBC on the basis of cytoplasm color.
Step
1: The coordinates of cytoplasm region is obtained by applying logical ExOR operation in between dilated region and region.
2: Find out the average Hue value of the cytoplasm region using the Eq. (9). The coordinates of pixels representing the cytoplasm region are obtained from the previous step. $H(i,j) = \begin{cases} \theta, & \text{if } B(i,j) \le G(i,j) \\ 360^\circ - \theta, & \text{otherwise.} \end{cases}$
(9) where $\theta = Cos^{-1}\{\frac{\frac{1}{2}[(R(i,j)-G(i,j))+(R(i,j)-B(i,j))]}{\sqrt{[(R(i,j)-G(i,j))^2+(R(i,j)-B(i,j))(G(i,j)-B(i,j))]}}\}$
H(i, j) denotes Hue value of *i*th row and *j*th column. $R(i,j)$, $G(i,j)$, $B(i,j)$ represents the red green, and blue components of the pixel (i,j)
3: Apply fuzzy based classifications, discussed in Algorithm 6, to classify the WBC type from average hue value. The fuzzy membership functions are built using training data set.
4: Return the decision with confidence value. maximum of the membership values
$Confidence = \frac{\text{maximum of the membership values.}}{\text{sum of the membership values.}}$
5: Stop

2.4.4. Texture feature based classification

Texture analysis is based on statistical properties of the histogram of intensity values [37]. It is applied on the selected regions of I3 matrix (discussed earlier). In texture-based classification, texture features, namely, smoothness, skewness and entropy are used [38]. The texture-based classification is suitable for distinguishing Basophil and Gametocyte stages of *Plasmodium vivax*. Three fuzzy membership functions for Basophil (f_B), Gametocyte stage of *Plasmodium vivax* [5] (f_P) and others (f_O) are built to classify them. The texture features Smoothness (S), Skewness (S_k) and Entropy (e) are defined in Eqs. (10), (11), (12).

$$S = 1 - \frac{1}{\left(1 + \sum_{i=0}^{L-1}(z_i - m)^2 P(z_i)\right)} \tag{10}$$

$$S_k = \sum_{i=0}^{L-1}(z_i - m)^3 P(z_i) \tag{11}$$

$$e = \sum_{i=0}^{L-1} P(z_i)\log_2 P(z_i) \tag{12}$$

where $P(z_i)$ indicates the number of pixels at intensity level z_i, L is the number of possible intensity levels and m is the mean intensity.

These three features are used to build each of the three fuzzy membership functions mentioned above. Each of the membership functions is defined in a form of the cubic polynomial.

$$\begin{aligned} f(S, S_k, e) = {} & c_1 + c_2 * s + c_3 * s^2 + c_4 * s^3 + c_5 * s_k + c_6 * s_k^2 \\ & + c_7 * s_k^3 + c_8 * e + c_9 * e^2 + c_{10} * e^3 + c_{11} * s^2 * s_k \\ & + c_{12} * s^2 * e + c_{13} * s_k^2 * e + c_{14} * s_k^2 * s + c_{15} * e^2 * s \\ & + c_{16} * e^2 * s_k + c_{17} * s * s_k * e + c_{18} * s * s_k \\ & + c_{19} * s * e + c_{20} * s_k * e \end{aligned} \tag{13}$$

where c_1, c_2, c_3, c_4, c_5, ..., c_{20} are coefficients. Values of the coefficients for each class are determined by forming 20 linear simultaneous equations using 20 randomly selected instances from the corresponding training set containing values of s, s_k, e and corresponding responses. The response is the percentage of belongingness to a particular class corresponding to a set of feature values of s, s_k and e. This set of 20 linear simultaneous equations is solved by using Gauss–Seidel iterative method. For each set of feature values, the fuzzy membership functions f_B, f_P and f_O are evaluated and classification is made on the basis of maximum membership value.

2.4.5. Final classification

In the present work, four independent classification approaches, based on four different feature sets, are used. For better classification, all these results are recombined for final classification. The regions are grouped into 8 categories viz Neutrophil, Lymphocyte, Monocyte, Gametocyte stage of *Plasmodium vivax*, Gametocyte stage of *Plasmodium falciparum*, Eosinophil, Basophil, and artefact. These categories are denoted by G1, G2, G3, G4, G5, G6, G7 and G8. Random forest machine learning technique is used for multi-stage classification as described in Algorithm 8. The term confidence is used to denote the strength of a decision. All individual classifications are based on selection of fuzzy membership function producing largest membership value. So, confidence is the ratio between maximum value from all fuzzy membership functions and aggregate value of all fuzzy membership functions.

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

Algorithm 8. To make final classification decision.

bf Require: All classification decisions along with confidence values. The confidence value is the belongingness (in percentage) of a region to a particular class.
bf Output: Final classification of the region. It classifies the types of WBCs, artefacts and parasites
bf Step
1: The detected regions are classified into Neutrophil, Lymphocyte, Monocyte, gametocyte stage of *Plasmodium vivax*, gametocyte stage of *Plasmodium falciparum*, Eosinophil, Basophil, and artefact. These classes are denoted by G1, G2, G3, G4, G5, G6, G7 and G8 respectively.
2: If the decision from shape feature is either G2 or G3 with high confidence, denoted as an index in Algorithm 5. Then apply Random forest classifier, which has been trained with area values of G2 and G3 categories upon the area value of the selected region. If it returns true, then the decision is G2 or G3 accordingly. Else apply Random forest classifier, which is trained with the area value and texture feature of the G4 category, upon the area value and texture of the region. If it is true, then the decision is G4 else it is G8. The G4 category has similarity in shape with G2 or G3 category, but G4 is small in size with a different texture.
3: If the decision from shape feature is G1. Then apply Random forest classifier, which has been trained with size_ratio and area value of G1 category. If it returns true, then consider the color feature to categorize them into G1, G6, G7 category; and then decision is taken accordingly. In G1, G6, G7 has similar size_ratio but cytoplasm color is different. If Random forest classifier returns false, then the category will be G8.
4: If the decisions from shape feature is any category with low confidence and size_ratio and also the texture feature categorizes the same with low confidence. Then the decision is G8.
5: If the decision from shape feature, size_ratio, texture feature are different from each other with similar conference level, then the decision is G8.
6: If the decision from shape feature is G5 with medium to high confidence and size_ratio and size also supports the decision, then the decision is G5.
7: Stop

Algorithm 8 provides better results than an individual decision. For example shape and size features are not very efficient to distinguish between G1 and G6. The color feature plays an important role to finalize the decision.

2.5. Report generation

WBC estimation procedure is divided into two parts: the first one is total WBC count, and another one is differential WBC count [4]. Total WBC count measures the total number of leukocyte in 1 μl of blood. The estimation of "Total WBC count" is defined as

$$\text{Total Count} = \left(\frac{\text{Total number of leukocyte found in 40 fields with 1000} \times \text{magnification}}{20}\right) * 30,000$$

Differential count of WBC measures the percentage of different types of WBCs in the blood sample. The differential count is crucial for determining the type of infection in the body. The Differential count is different for different age groups and genders [1,9]. Differential counting process continues until 100 WBCs are counted and classified. On the basis of this data, the Differential count result is made.

2.6. Time complexity analysis for the algorithms

The input images are 960×1280. So the total number of pixels is 1,228,800 and it is denoted by n. Each steps from step 1 to 7, excluding step 4, in Algorithm 1 takes n number of iterations. Eq. (2) in step 4 requires $L \times L$ number of iterations. The value of L is always 256 (maximum intensity level). So the value of $L \times L$ is very less compared to n. Hence the time complexity of Algorithm 1 is in the order of $O(n)$.

In Algorithm 2 steps 1, 2 and 5 requires n numbers of iterations. 10% of total pixels are supplied to step 3 as input, so number of iteration for step 3 is $n/10$. Step 4 calls step 3 until there is no change between the old and new threshold value. On an average step 4 iterates not more than 30 times. So the time complexity of Algorithm 2 is also $O(n)$.

Consider the case of Algorithm 3, step 2 and 3 are executed P times. P is the number of separate regions which may contain WBC in the input binary image. Both the steps 2 and 3 require K internal iterations. Where K is the average number of pixels used to represent a separate region. The total number of iterations is $P \times K$. This must be less than the total number of pixels (n). In real situation, value of P is less than 10 and K is 3500, on an average. The time complexity of Algorithm 3 is restricted within $O(n)$.

Step 1 of Algorithm 4 has n number of iterations. Step 2 to 4 are repeated for T times. For execution of Step 2 to 4 have K internal iterations. So $T \times K$ iterations are required for each step. The value of T varies from case to case but it never goes beyond 50. So value of $T \times K$ is not more than n. Hence the complexity of Algorithm 4 is $O(n)$.

In Algorithm 5, step 3 describes a sorting procedure and step 6 and 10 implements Eq. (5). Each of this three steps has B^*B iterations. B is the number of boundary points for a single region. Average value of B is 700. So value of B is less than n it can be said that the time complexity of Algorithm 5 is within $O(n)$.

Each step of Algorithm 6, except step 6, has K number of iterations. In step 6, Eq. (8) is calculated which requires $I \times I$ number of iterations. Where I is the total number of training sample. $I \times I$ is less than n this, complexity of Algorithm 6 does not exceed $O(n)$.

All the steps of Algorithm 7 are straight forward and time complexity is restricted within $O(n)$.

Algorithm 8 uses random forest classifier, which is trained with I numbers of training sets. Once the training of the classifier is completed it requires constant time to execute. Because there is no internal iteration for the steps of Algorithm 8. Hence the complexity of this algorithm is $O(1)$, if we do not consider the time to build the internal data structure of random forest classifier (which is build only once).

3. Results and discussion

The sample blood smear slides along with manual reports are collected from Kolkata Municipal Corporation (KMC) health centers, Kolkata, India. One of the authors has the permission to work with KMC pathology labs. The authors and their respective institutes are not directly involved with blood sample collection, smear slide preparation, etc. KMC health centers use Leishman stain to color the slides. Some of the sample slides contain *Plasmodium vivax*. The stained blood samples are examined under 1000× magnification. To calculate total WBC count for a single blood sample, 40 fields are examined. For field selection, only those fields are considered where 50% of RBCs are partially overlapped and rest 50% are not overlapped [35]. The WBC estimation fully depends on the RBC distribution. For differential count, the number of fields to be examined is not fixed. Field examining process stops when 100 WBC are detected and classified.

Table 1 shows the comparative study of the proposed system, conventional microscope-based manual diagnosis system, Bergen et al. method [12] (method 1) and Dorini et al. method [13] (method 2).

150 slides, along with reports, are used to test the system. The system takes 100 images from each of the slides, to calculate the

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

Table 1
Comparison of the proposed method with other methods.

Metric and parameters	Proposed system total count	Manual system total count	Method-1 total count	Method-2 total count
False positive	20	31	85	87
True negative	67	56	2	0
True positive	464	444	476	480
False negative	23	43	11	7
Error rate	7.5%	12.9%	16.7%	16.3%
Precision	95.8%	93.4%	84.8%	84.6%
Sensitivity	95.2%	91.1%	97.7%	99.3%
Specificity	75.8%	64.3%	2.2%	0%

Table 2
The confusion matrix of the proposed system.

		Predicted class Neutrophil	Eosinophil	Basophil	Monocyte	Lymphocyte	Non-WBC
Actual class	Neutrophils	201	17	12	3	2	7
	Eosinophil	2	20	1	4	0	3
	Basophil	4	2	18	0	0	4
	Monocyte	3	1	2	38	1	3
	Lymphocyte	4	2	1	1	125	6
	Non-WBC	5	2	8	3	2	67

differential count and first 40 images are used for the total count. So, the total number of images captured is $150 \times 100 = 15,000$. Out of these 15,000 images, only 300 images are selected to train the system (to build the fuzzy membership functions and train the random forest classifiers). Among these 150 slides, 10 contain *Plasmodium vivax*. Methods proposed by Bergen et al. and Dorini et al. are not able to distinguish between WBC and *Plasmodium vivax*, so they produce erroneous result. The comparison was made on 400 images where 487 WBC, 31 *Plasmodium vivax* and 56 artefacts are present. Out of 487 WBC, the differential count of WBC are as follows Neutrophil 242, Eosinophil 30, Basophil 28, Monocyte 48 and Lymphocyte 139.

For the proposed system, false positive in the total count is 20 and this is because sometimes confusion arises between artefacts and malarial parasites. Artefacts and malarial parasite are detected mainly on the basis of texture features, which partly depend on microscope focusing. Poor focusing of microscope reduces fine variations in the texture feature values, and confusing decisions are generated. False positive of a manual system is higher than the proposed system. It varies from person to person; the false positive for the total count for this dataset is 31. There is no mechanism in method-1 and method-2 to distinguish between WBC, and other Leishman stained blood components like *Plasmodium vivax* and artefacts. For this reason, false positive is quite high for this dataset.

False negative in total count for proposed system is 23. Sometimes Neutrophil nucleus is fragmented into separate parts, which are not recombined in false contour elimination stage. The small parts are sometimes eliminated or misclassified. This is the reason of false negatives. The number of false negative for the manual system is 43 (may be due to overlook or human Boredom). Number of false negative is negligible for method-1 and method-2 because they detect all Leishman stained objects with considerable size.

True positives for proposed system is $(487 - 23) = 464$. In case of manual system true positives for total count are 444, whereas those for method-1 and method-2 are 476 and 480 respectively.

True negatives in total count for the proposed system is $(87 - 20) = 67$. In case of manual system true negatives in total count are 56, and those for method-1 and method-2 are 2 and 0 respectively.

Small premature WBCs are sometimes classified as the malaria parasite. This is one of the causes of false negatives. On the other hand overlapped platelets are considered as a premature WBC, it is an example of false positives.

Table 2 shows confusion matrix of the proposed system. Bergen et al. method and Dorini et al. methods detects only WBC regions. The individual classification of WBC is not focussed in their work. So performance analysis regarding the classification of individual WBC type (differential count) is not applicable to these methods. In case of manual counting, the performance analysis regarding the classification of individual WBC type solely depends on work load. The same image may be classified properly or misclassified; it depends solely on the concentration of mind. So, the result may be contradictory and for this reason the confusion matrix of the manual system is not calculated. In Table 2, six classes are considered; where *Plasmodium vivax* and artefacts are grouped into Non-WBC class. The six classes are Neutrophils, Eosinophils, Basophils, Monocytes, Lymphocytes and Non-WBC. The texture of Basophils is quite similar with *Plasmodium vivax* so sometimes Basophils are misclassified as Non-WBC.

For the proposed method, Table 3 shows the performance of differential count where individual WBC types are classified.

For the proposed system, the sensitivity of the classification of individual type of WBC (differential count) differs from the sensitivity of the aggregate WBC detection (total count), as shown in Table 2. The reason for this difference is that Neutrophils, and Eosinophils look very similar to each other. This increases the chances of misclassifications, which reduces the sensitivity of Neutrophils, and Eosinophils. The look alike issue also decreases the sensitivity of Basophils, Monocytes, Lymphocytes. All this misclassification issues are within WBCs, so this misclassification problem

Table 3
The performance of individual WBC classification using proposed method.

Metric and parameters	Neutrophil	Eosinophil	Basophil	Monocyte	Lymphocyte	Non-WBC
False positive	18	24	24	11	5	23
True negative	314	520	522	525	430	464
True positive	201	20	18	28	125	67
False negative	41	10	10	10	14	20
Sensitivity	83.0%	66.6%	64.2%	73.6%	89.9%	77.o%
Specificity	94.5%	95.5%	95.6%	97.9%	98.8%	95.2%

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038

does not affect the WBC and Non WBC classification (total count). Finally, the sensitivity of the aggregate diagnosis is satisfactory. On the other hand, Table 1 shows a part of Non-WBC regions as True negative, where as Table 3 shows a part of Non-Neutrophils region (consider the first column of Table 3), which has higher value than Non-WBC regions. A higher true negative value increases Specificity.

4. Conclusion

An automatic hybrid system based on Fuzzy logic and Random forest has been proposed for calculating the total count and a differential count of WBCs in blood samples. The microscopic images of Leishman stained blood smear slides are taken as input, and a report is generated mentioning the total count and a differential count of WBCs in blood samples. The testing cost per patient is as low as a manual system, and accuracy is comparable with or sometimes better than that of the manual systems. Minimum technical skill is required to prepare slides, and almost no technical skill is needed to operate the system. Hence, it can be deployed at remote place of any developing country with minimum infrastructure. Also, this automatic system is capable to reducing the huge work pressure upon the government funded public health centers in the developing countries. The performance of this system can be improved further by using images captured with fluorescence microscopes, which will help to reduce artefacts [39] in the captured images. As a succeeding phase of work, authors are planning to introduce GPS [40] to design the region wise and year wise complete disease map. This map will be helpful for taking precautionary measures in the forthcoming years. Finally, this system has the potential to become a powerful tool for telemedicine system.

Acknowledgements

Authors are thankful to Dr. Soumendu Datta and Dr. Abhijit Sen; for providing relevant knowledge regarding WBC counting and Kolkata Municipal Corporation for providing sample slides during the progress of the work. Authors are grateful to “Department of Biotechnology GOI funded Telemedicine Compatible Computerized Hemocytometer Analyzer for Dengue Detection and Treatment project” for providing infrastructural facilities during the progress of the work.

References

[1] R. Drake, A.W. Vogl, A.W. Mitchell, Gray’s Anatomy for Students, Elsevier Health Sciences, 2014.
[2] F.R. Sabin, Studies of Living Human Blood-cells, Johns Hopkins Hospital Bulletin, 1923.
[3] M.M. Wintrobe, J.P. Greer, Wintrobe’s Clinical Hematology, vol. 1, Lippincott Williams & Wilkins, 2009.
[4] H.K. Walker, W.D. Hall, J.W. Hurst, Clinical Methods: The History, Physical, and Laboratory Examinations, 3rd ed., Butterworths, 1990.
[5] World Health Organization and Center for Disease Control, Basic Malaria Microscopy: Tutor’s Guide, World Health Organization, 2010.
[6] T.H. Brannagan, Y. Zhou, HIV-associated Guillain-Barré syndrome, J. Neurol. Sci. 208 (1) (2003) 39–42.
[7] D.M. Jaffe, G.R. Fleisher, Temperature and total white blood cell count as indicators of bacteremia, Pediatrics 87 (5) (1991) 670–674.
[8] L.O. De Labry, E.W. Campion, R.J. Glynn, P.S. Vokonas, White blood cell count as a predictor of mortality: results over 18 years from the normative aging study, J. Clin. Epidemiol. 43 (2) (1990) 153–157.
[9] N.H. Rasmussen, L.N. Rasmussen, Predictive value of white blood cell count and differential cell count to bacterial infections in children, Acta Paediatr. 71 (5) (1982) 775–778.
[10] R.H. Grimm, J.D. Neaton, W. Ludwig, Prognostic importance of the white blood cell count for coronary, cancer, and all-cause mortality, J. Am. Med. Assoc. 254 (14) (1985) 1932–1937.
[11] K. Jiang, Q.M. Liao, Y. Xiong, A novel white blood cell segmentation scheme based on feature space clustering, Soft Comput. 10 (1) (2006) 12–19.
[12] T. Bergen, D. Steckhan, T. Wittenberg, T. Zerfass, Segmentation of leukocytes and erythrocytes in blood smear images, in: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, IEEE, 2008, pp. 3075–3078.
[13] L.B. Dorini, R. Minetto, N.J. Leite, Semiautomatic white blood cell segmentation based on multiscale analysis, IEEE J. Biomed. Health Inf. 17 (1) (2013) 250–256.
[14] H. Yamamoto, M. Oka, Automatic blood analyzer, US Patent 4,030,888 (June 21, 1977).
[15] S. Weissenbacher, B. Riond, R. Hofmann-Lehmann, H. Lutz, Evaluation of a novel haematology analyser for use with feline blood, Vet. J. 187 (3) (2011) 381–387.
[16] Olympus Opto Systems India Pvt. Ltd., Olympus UIS2 Infinity Optics, rev. 3 (4 2014).
[17] P. Ghosh, D. Bhattacharjee, M. Nasipuri, D.K. Basu, Medical aid for automatic detection of malaria, in: Computer Information Systems-Analysis and Technologies, Springer, 2011, pp. 170–178.
[18] TOSHIBA [Toshiba Semiconductor], PWM Chopper-Type bipolar Stepping Motor Driver IC, rev. 2 (5 2006).
[19] M. Khalilian, A. Abedi, A.D. Zadeh, Position control of hybrid stepper motor using brain emotional controller, Energy Proc. 14 (2012) 1998–2004.
[20] S. Bhattacharyya, A brief survey of color image preprocessing and segmentation techniques, J. Pattern Recognit. Res. 1 (1) (2011) 120–129.
[21] S. Bhattacharyya, U. Maulik, Soft Computing for Image and Multimedia Data Processing, Springer, 2013.
[22] S. Bhattacharyya, U. Maulik, P. Dutta, Multilevel image segmentation with adaptive image context based thresholding, Appl. Soft Comput. 11 (1) (2011) 946–962.
[23] S. Bhattacharyya, P. Pal, S. Bhowmick, Binary image denoising using a quantum multilayer self organizing neural network, Appl. Soft Comput. 24 (2014) 717–729.
[24] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Pearson Education India, 2009.
[25] D.J. Kang, A fast and stable snake algorithm for medical images, Pattern Recognit. Lett. 20 (5) (1999) 507–512.
[26] R. Adams, L. Bischof, Seeded region growing, IEEE Trans. Pattern Anal. Mach. Intell. 16 (6) (1994) 641–647.
[27] A. Stalder, G. Kulik, D. Sage, L. Barbieri, P. Hoffmann, A snake-based approach to accurate determination of both contact points and contact angles, Colloids Surf. A: Physicochem. Eng. Asp. 286 (1) (2006) 92–103.
[28] M. Nixon, Feature Extraction & Image Processing, Academic Press, 2008.
[29] Z.Q. Hong, Algebraic feature extraction of image for recognition, Pattern Recognit. 24 (3) (1991) 211–219.
[30] D.J. Hurley, M.S. Nixon, J.N. Carter, Force field energy functionals for image feature extraction, Image Vis. Comput. 20 (5) (2002) 311–317.
[31] M. Desbrun, E. Kanso, Y. Tong, Discrete differential forms for computational modeling, in: Discrete Differential Geometry, Springer, 2008, pp. 287–324.
[32] P. Ghosh, D. Bhattacharjee, M. Nasipuri, D.K. Basu, Automatic white blood cell measuring aid for medical diagnosis, in: International Conference on Process Automation, Control and Computing (PACC), IEEE, 2011, pp. 1–6.
[33] A. Yardimci, Soft computing in medicine, Appl. Soft Comput. 9 (3) (2009) 1029–1043.
[34] O. Cordón, F. Herrera, Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems, Fuzzy Sets Syst. 118 (2) (2001) 235–255.
[35] Y. Purwar, S.L. Shah, G. Clarke, A. Almugairi, A. Muehlenbachs, Automated and unsupervised detection of malarial parasites in microscopic images, Malaria J. 10 (1) (2011) 364.
[36] I. Singh, Textbook of Human Histology: With Colour Atlas & Practical Guide), Jaypee Brothers Publishers, 2011.
[37] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1) (1996) 51–59.
[38] A. Materka, M. Strzelecki, Texture Analysis Methods – A Review, COST B11 Report, Brussels, Technical University of Lodz, Institute of Electronics, 1998, pp. 9–11.
[39] F. Kawamoto, Rapid diagnosis of malaria by fluorescence microscopy with light microscope and interference filter, Lancet 337 (8735) (1991) 200–202.
[40] E. Kaplan, C. Hegarty, Understanding GPS: Principles and Applications, Artech House, 2005.

Please cite this article in press as: P. Ghosh, et al., Blood smear analyzer for white blood cell counting: A hybrid microscopic image analyzing technique, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2015.12.038