International Journal of Engineering & Technology, 7 (4) (2018) 3642-3646

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.20072
Research paper



White blood cell recognition via geometric features and naïve bays classifier

Hasan A. Kazum [1] *, Faisel G. Mohammed [2]

[1] Department of computer science. College of science, Baghdad University, Baghdad, Iraq
[2] Department of Remote Sensing and GIS, College of science, Baghdad University Baghdad, Iraq
**Corresponding author E-mail: hassanakazum@yahoo.com*

Abstract

Blood assessments are of the maximum crucial and frequently asked medical examinations. A manual microscopic evaluation must be done while a blood pattern is suspicious of abnormality. This manual technique is tedious, time ingesting and subjective. Automating microscopic blood type is appropriate to assist the pathologists to hurry-up and induce the consequences accuracy.
Segmentation is the primary and common step in computerized WBCs category. On this paper, have been presented a powerful method for automated WBCs nuclei segmentation. The technique is based on gray scale contrast enhancement and then using Otsu thresholding tech-nique to segment WBCs. There are four features have found to extract the data from the segmented image. These features are (Area, Perim-eter, diameter, Circularity. Then these data was classified using Naive Bayes classifier under weka program. The approach is examined on 260 blood pictures. The class overall performance is quantitatively evaluated at the take a look at set to be 97,1 %. This overall performance is excessive in comparison to other related work done at the identical dataset.

Keywords: *WBC; Naïve Bayes; Weka; Classification; Segmentation.*

1. Introduction

Visual analysis of two-dimensional (2D) pathology images provides quantitative information about the presence and absence of disease process and helps diagnosis of disease [1]. In automated diagnosis systems, though human intervention can be necessary, it is desirable that the amount of intervention is minimum. Moreover, researches have shown that with improved segmentation accuracy, better diagnosis performance can also be achieved [2].
To a certain extent, the performance(of an automatic white blood(cell classification system(depends on(a good(segmentation algorithm for segmenting(white blood cells from their background[3]. There are many different approaches (e.g., clustering, Schmidt(orthogonaliztion method,(edge detection, region(growing, watershed, colors , and(support vector machine (SVM) to segment white(blood cells(from the(background [4].
Each approach has its advantages(and disadvantages. For example, the conventional color-based methods and the thresholding method are simple but are not able to accurately segment the white blood cells from the background[5]. Some approaches (e.g., the SVM method and the region growing method) can provide reasonably accurate segmentation results, but they are either costly to be implemented or require high computational resources [6]. A review on some of the general segmentation methods can be found in. While some color-based segmentation methods were directly conducted on the RGB color space, some approaches adopted the color space (especially on the component) [7].
WEKA (Waikaato Environment for Knowledge Acquition) is the most widely used data mining tool which support huge amount of data mining algorithm for classification. The WEKA software(was developed in the University of New Zealand. A number(of data mining methods are implemented(in the WEKA software. Some of them(are based on decision trees(like the J48 decision(tree, some are(rule-based like(ZeroR and decision tables, and some(of them are based on probability and regression, like the(Naïve Bayes(algorithm[8].
In this paper, we developed a system for implementing an automatic white blood cell classification system. First of all, we try to identify the color characteristics of the pixels of the nucleus and granule of cytoplasm of white blood cells in the color space. Based on the found discriminating region and a morphological process, we can segment a white blood cell from a smear image. In the following, we extract Four kinds of features (i.e., Area, Perimeter, dimeter, circularity) from the segmented cell region. These features are fed into Naïve Bayes classifier under weka for classifying five(types of the(white blood(cells.

2. Methodology

The current research work, leishman stain approach was used for staining samples of blood smear pictures all of the WBCs have been stained in blue shade, and the dimensions of pics is (640×480) [9]. Picture data of that microscopic blood picture have been obtained from local hospital in Iraq and websits. This studies paintings is executed at the MATLAB R2014a picture processing toolbox. In figure (1) firstly, photo preprocessing implemented to the input picture as achieved , in which the nucleus region accompanied via Otsu's thresholding segmentation and morphological operations [10].

Copyright © 2018 Hasan A. Kazum, Faisel G. Mohammed. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3. Proposed methodology

In this study, the segmentation of WBCs was done based on the mathematical operations, thresholding approach and mathematical morphing . Its applied to obtain smoothing image, then followed with classification of cell using of Naïve Bayes classifier under weka.

3.1. Image pre-processing

Photo preprocessing may be very critical in clinical analysis so that it will get excessive great medical picture, but many factors have effect on its acquisition despite the fact that photo processing cannot offer new facts for prognosis. It can enhance the visible effect to diagnose appropriately; so that the consequent picture is higher suited for machine interpretation. As seen in Figure (1) This degree consists of the authentic photo conversion which the colour pics (RGB) are transformed to the gray scale. At this step, our consciousness is at the mathematics operation carried out to the photograph a good way to be easily segmented; then vert(the input image to(grayscale, then make two copies(of gray-scale(photo.In a single(replica histogram(equalization H(i,j) and contrast(starching C(i, j) [11].

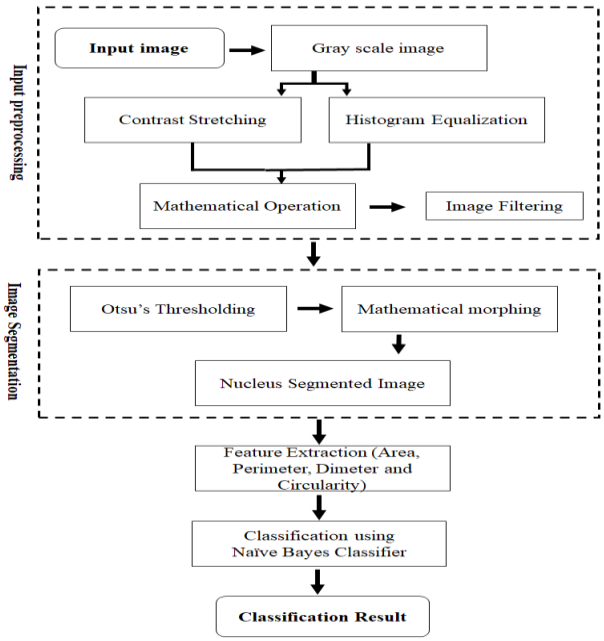


Fig. 1: Block Diagram of the System.

Then,imathematical operations likeiaddition are applied on the twoiimages C (i,j) and H (i,j), and the output is image R1(i,j) as shown in equation (1) which highlight nucleusiof leukocytesiand brightens all otheriblood componentsiimage, and then imageisubtraction is doneiwith R1 (i,j) and H (i,j), which highlighted all of the objects in image. The output ofisubtraction operation is image R2(i,j) as shown in equation (2)

$$R1(i,j) = C(i,j) + H(i,j) \quad (1)$$

$$R2(i,j) = R1(i,j) - H(i,j) \quad (2)$$

Finally, combiningiboth the images R1(i,j) and R2(i,j) to get image R3(i,j) as shown in equation(3), that resultsiminimum effectiof distortioniin nucleus

$$R3(i,j) = R1(i,j) + R2(i,j) \quad (3)$$

3.2. Segmentation using Otsu's thresholding

Inithis stage, WBC images are segmented toiproduce a numberiof regions; each regionirepresents one cellifrom smear of blood. Thresholding process is the handiest approach for segmenting specific picture. It's foremost purpose to partition all pixels of picture into foreground and the historic beyond primarily based on intensity of gray or textures level. We've distinctive typeiofithresholding strategies together with international, variable and more than one thresholding. In the global thresholding, the suitable threshold fee T is ready for the complete picture and on idea of photo[12].

$$I_{bin}(i,j) = \begin{cases} 1\, I(x,y) \geq T \\ 0\, \text{ohterwise} \end{cases}$$

Thresholdivalue T changeover image relies upon on whether or not local or adaptive thresholding. In localithresholding, T relies upon a community of each pixel (x,y). Meanwhile, adaptive thresholding, T is a feature of pixel (x, y). A couple of thresholding targets to discover more than one threshold values to split multiple items.

$$I_{bin}(i,j) = \begin{cases} a, \text{if } I_{gray}(x,y) > T_2 \\ b, \text{if } I_{gray}(x,y) < T_2 \\ c, \text{if } I_{gray}(x,y) \leq T_2 \end{cases}$$

In the current research work, Otsu's global thresholding approach was used, which it is limiting the weightediwithin-elegance varianceiofithe thresholder foregroundiand historical past pixelsiand operateiat gray-level bimodularihistogram or picture to establish a choicest threshold T, then used to binaries the photo [16].
The weightediof inside-elegance variance is given by

$$\sigma_w^2(T) = v_1(T)\sigma_1^2(T) + v_2(T)\sigma_2^2(T)$$

Where σ_w is the within class variance, v1 is the class variance of foreground, and v2 background, class probabilities are estimated from histogram as:

$$V_1(T) = \sum_{i=1}^{T} p(i)\; V_2(T) = \sum_{i=T+1}^{L} p(i)$$

WhereP(i) is the frequency i, L is the quantity of gray in picture. The elegance manner of foreground and historical past are estimated through

$$\mu_1(T) = \sum_{i=1}^{T} \frac{iP(i)}{q_1} \quad \mu_2(T) = \sum_{i=T+1}^{L} \frac{iP(i)}{q_2}$$

The foreground variance and historical past pixels are given through

$$\sigma_2^2(T) = \frac{1}{v_2(T)} \sum_{i=1}^{T} (i-\mu_2)^2\, P(i)$$

$$\sigma_2^2(T) = \frac{1}{v_2(T)} \sum_{i=T+1}^{L} (i-\mu_2)^2\, P(i)$$

Now, we forestall and discover the vaule of T from (1,256), in order that weighted sum of within-elegance variance is minimal.

3.3. Mathematical morphology operations

Photo processing is continuously offering ienhancement, segmentation, convexihull, recuperation, facetidetection, texture analysis, shape, and size analysis [13]. Mathematical morphological operations arei nonlinear, translation invariant adjustments photo analysis method whichiextracts objects from image through describing theirigeometric structure. As we keep in mind only binaryiimages so right here we handiest supply detail of binary morphologicalioperations. In ouriresearch, we focus on the photograph morphingias it gets rid of theiunwanted gadgets like RBCs and platelets from the photograph. In mostiof the studies work mathematicali-

morphing used because the very last step to clean the vicinity ofihobby. There are four critical operators of mathematicalimorphing,ierosion, dilation,istarting, and ultimate. The 2 primaryioperators or mathematicalimorphing are erosioniand dilation, and at the mixture of them, other operators may be shaped[14]. Supposeiwe have photograph r andistructuring element s. Then the operations are denoted as:

$$Erosion\ R \ominus S = \bigcup_{s \in S^{R-s}}$$

$$Dilation\ R \oplus S = \bigcup_{r \in R^{Sr}}$$

Commencing and the closingimorphology operations areiderived from the mixtureiof abrasion (erosion) and theidilation. Beginning used toismoothest the contour of gadgets; breaks slim isthmuses, and eliminateiskinny protrusions through doing awayiwith small objectsifrom foreground. Closing isilikewise used to easy contoursithrough doing away with the smalliholes, fusesinarrow breaks, long skinnyigulfs, and fill theigaps inicontour.

$$Opening\ R \circ S = \bigcup_{S \subseteq R} B_x$$

$$Closing\ R \cdot S = \bigcup_{S \subseteq R} B_x$$

3.4. Feature extraction

Afterithe pictures areisegmented iniprevision level, the popularity of segmented items or areas from different gadgets is wanted. Consequently, every item is taken into consideration as pattern with its functions, the capabilities are extracted to lessen the unique information set and to differentiate enter sample from some other. The characteristic is decided which depend most effective at the characteristic encompass area, perimeter, circularity. For acquiring those functions, first, crop the nucleus from the entire image so that features can be extracted of only that region which is required. Area, perimeter, diameter, and Circularity of each segmented nucleus calculated. The feature extraction can be done using Geometrical Features[15].

3.5. Classification via naïve bays under weka program

The WBCs imagesidatabase (training set) containsiattribute valuesirepresented with attributes (geometricifeatures). This attributesiare entered to theiclassifier modelifor learning (training) phase. The predictioniof newicases depends on the builticlassifier in learning (training) phase; all the WBCsiimages used in prediction (testing) phaseiareinew WBCs images that do notiexist in the learning (training) phase. Theiprediction (testing) phase is used to test the WBCsiimages (testing set). NaïveiBayesiclassifier is used for leukocyte classification that's a primarilyibased on statistical[16]. It makes useiof Bayes Theoremiand based is totally in easy probabilisticiclassifier with impartial Naïveiassumptions manner that the value of the features is unbiasediof the lifestyles or non-life of another features. Every of those features makes contributionsiindependently to the opportunity First, calculate the mean and the varianceifrom each feature of every elegance cj as unbiased variables are assumed after which save them, after that, we have counted the possibility of earlier p(c) ,
The basis of naive Bayesian classification is Bayes' rule, shown in (2.4), which allows the computation of a conditional probability, P(b|a), given the two class priors P(a) and P(b), and the conditional probability of P(a|b).

$$P(b|a) = \frac{P(a|b)\,P(b)}{P(a)}$$

The method requires that the probability of each attribute value ai relative to the class be known or estimated, and employs the product rule—that is, assumes conditional independence amongst the attribute values P(ai |c)—resulting in the following formulation:

$$P(c|X) = \frac{p(c)\prod iP(ai|c)}{P(X)}$$

For a NB classifier to make a prediction, the probability P(c|X) for all possible classifications, c ∈ C, is calculated in order to find the maximum-likelihood classification hypothesis, often referred to as maximum a posteriori or MAP. For this computation, the P(X) terms are constant and therefore typically ignored. Thus the prediction task is reduced to computing Equation 2.6 for each instance in the test set.

$$PredX(c) = argmax\ c\ P(c)\prod_i P(ai|c)$$

The magnificence which has a most posterior chance, white blood cellular belongs to that magnificence. On this studies, the trouble could be very correctly solved by way of the Naïve Bayes classifier and offers us properly consequences[17].

4. Results and dissuasion

A regular peripheral blood sample changed into taken andistained. Regularilight microscope turned into used toiacquire virtual pix from theiblood slideiusing a 100x. An analog rate-coupled device (CCD) shade camera is connectedito theimicroscope to capture shade photographs (i.e., 640x480 pixels). Total of (260) photographs had been used on this take a look at with the subsequent 5 cell kind's distributions.
The nucleus segmentation sequence of White blood cell has been shown in Figure (2). At the first step, after inputting the image, the original color of the image changed into the gray color, that leads to form two copies, one for contrast stretching and one for histogram equalizations after the mathematical operation, the preprocessing step is ended. Otsu thresholding and morphological operation were done to complete segmented the image and obtain cell image with segmented nucleus.

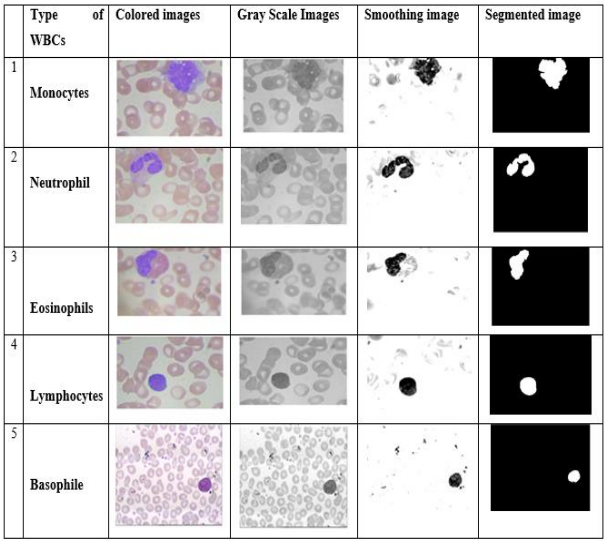


Fig. 2: The Stage of Segmentation Process.

The proposed algorithm is used geometric features to educate and check White blood cellular pictures, the consequences of making use of geometric features are proven in Tables (1). The geometric features which can be extracted from WBCs picture are area, perimeter, Diameter, and circularity. There may be a crucial step earlier than class technique, the numerical values acquired from features have to be transformed to specific values in order to be utilized in training the naïve Bayes classifier, which consists of dividing the features values in to periods of the values relying on the selected integer quantity. The outputs from this computation are utilized by the Naïve Bayes set of rules.
The experimentalievaluation of the current algorithm is performed initwo stages (training and testing). The mode 60% meanithat 60%

of images have been used for training phase, and 40% ofiimages have been used for testingiphase where (155) imagesiare selected foritraining phase and (105) imagesiare selected for testing phaseiand obtained 97,1%. Figure (3) visualized the dataset of mode 60% under weka program. Figure (4) showed the accuracy of each cell types of white blood cell under mode 60%.

Table 1: Accuracy Rate for 60%

WBC	No. of Training Image	No. of Testing Image	Recognized WBC		Classification Rate in %
			Yes	No	
Monocytes	10	8	8	0	100%
Neutrophil	91	63	62	1	98.4%
Eosinophils	16	11	9	2	81.8%
Lymphocytes	23	13	13	0	100%
Basophile	15	10	10	0	100%

```
=== Summary ===

Correctly Classified Instances         102               97.1154 %
Incorrectly Classified Instances         3                2.8846 %
Kappa statistic                          0.9515
Mean absolute error                      0.0324
Root mean squared error                  0.1154
Relative absolute error                 13.2806 %
Root relative squared error             33.4388 %
Total Number of Instances              105

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC
ROC Area  PRC Area  Class
                 0.984     0.000     1.000      0.984     0.992      0.980
0.998     0.999     neutro
                 0.818     0.011     0.900      0.818     0.857      0.842
0.936     0.873     eosi
                 1.000     0.011     0.929      1.000     0.963      0.958
0.997     0.982     lympho
                 1.000     0.010     0.800      1.000     0.889      0.890
1.000     1.000     mono
                 1.000     0.000     1.000      1.000     1.000      1.000
1.000     1.000     baso
Weighted Avg.    0.971     0.003     0.973      0.971     0.971      0.962
0.992     0.984

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 62  1  0  0  0 |  a = neutro
  0  9  1  1  0 |  b = eosi
  0  0 13  0  0 |  c = lympho
  0  0  0  8  0 |  d = mono
  0  0  0  0 10 |  e = baso
```

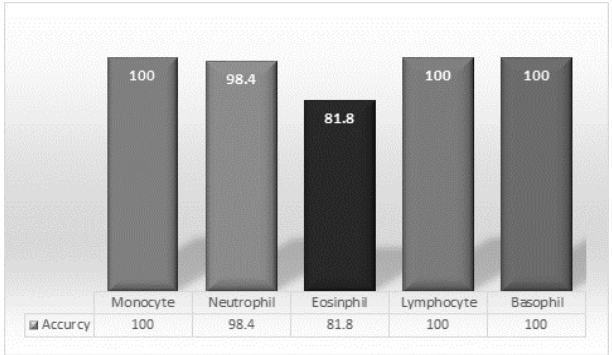


Fig. 3: Naïve Byes Classifier Results under Weka with Mode of 60%.

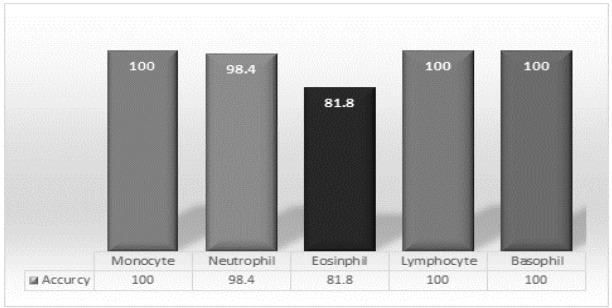


Fig. 4: Chart of the Accuracy of WBCS Types under Weka with Mode of 60%.

The confusion matrix of proposed set of rules overall performance has been acquired from the testing component through the usage of Geometric features may be proven in table (2). The confusion matrix ought to be examined as follows: rows suggest the item to apprehend, and columns suggest the label the classifiers buddies at this item.

Table 2: Confusion Matrix of 60% Mode Using Geometric Features

Class of WBCs	Neutro-phil	Eosino-phil	Lympho-cyte	Mono-cyte	Baso-phil
Neutrophil	62	1	0	0	0
Eosinophil	0	9	1	1	0
Lympho-cyte	0	0	13	0	0
Monocyte	0	0	0	8	0
Basophil	0	0	0	0	10

5. Performance comparison between Naïve Bayes classifiers and another classifiers

Finally, when comparing the results of WBCs classification with mode 60% using Naïve Bayes under weka with decision tree and SVM (Support vector machine) with the same dataset that used for Naïve Bayes classifier under weka with mode 60%. The Naïve Bayes have achieved the highest accuracy. On the other hand, SVM toolkit had the lowest accuracy measures that reached to (88.4615%). Meanwhile, Tree decision exhibited (95.1923%) accuracy when using mode of 60%. Figure (5) display the differential between the Naïve byes classifier, SVM and decision tree with mode 60%.

These results showed that Naïve byes classifier under weka program tool is better than the other to be used for a classification task, this is may be due to the kind of data sets used, or maybe there are some differences in the way the algorithms were implemented within the tools themselves. A NaiveiBayes classifier is a simple classifier[18]. However, althoughiit is simple, Naive Bayes can outperform more sophisticatediclassification methods. Besides that, it also has exhibitedihigh accuracy and speed when appliedito large database[19]. Moreover, it is very fast for both learningiand predicting. Its learning timeiis linear in the number of examples and its prediction time isiindependent of the number of examples[20]. Naïve Bayes classifier is alsoifast, consistent, easy to maintain and accurateiin the classification of attributeidata.

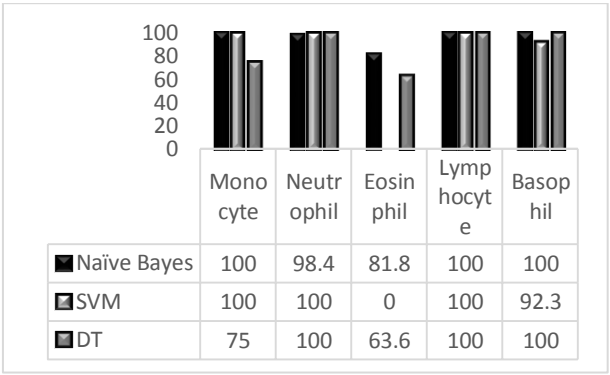


Fig. 5: Performance Comparison between Naïve Byes Classifiers, SVM and Decision Tree.

6. Conclusions

The possibilities of a final hit results of White blood cells commonly depend on its early detection and prognosis. This guide inspection technique is time-ingesting and mistakes susceptible. Accordingly, a pc-primarily based machine for computer-aided detection can also offer an assistive diagnostic device for pathologists. The automatic segregation and identity algorithm pursuits to lessen the latency duration concerned in a remedy, that's from time to time lifestyles threatening. The advanced computerized system is examined with Naïve Bayes Classifier at the dataset of (260) pretested samples; the accuracy carried out is 97,1 %. The consequences display that set of rules proposed achieves an appropriate overall performance for the White blood

cellular, similarly, the devised technique additionally addresses the segmentation of overlapping cells.

7. Disclosure statement

The authors declare that there are no conflicts of interest. The authors alone are responsible for the content and writing of the article.

References

[1] Razzak MI, Naz S, Zaib A: Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Classification in BioApps. Springer; 2018: 323-350. https://doi.org/10.1007/978-3-319-65981-7_12.

[2] Shafique S, Tehsin S: Computer-Aided Diagnosis of Acute Lymphoblastic Leukaemia. Computational and mathematical methods in medicine 2018, 2018.

[3] Liu Z, Liu J, Xiao X, Yuan H, Li X, and Chang J, Zheng C: Segmentation of white blood cells through nucleus mark watershed operations and mean shift clustering. Sensors 2015, 15(9):22561-22586.

[4] Habibzadeh M, Krzyżak A, Fevens T: White blood cell differential counts using convolutional neural networks for low-resolution images. In: International Conference on Artificial Intelligence and Soft Computing: 2013. Springer: 263-274. https://doi.org/10.1007/978-3-642-38610-7_25.

[5] Alférez Baquero ES: Methodology for automatic classification of atypical lymphoid cells from peripheral blood cell images. 2015.

[6] Dragozi E, Gitas IZ, Stavrakoudis DG, Theocharis JB: Burned area mapping using support vector machines and the FuzCoC feature selection method on VHR IKONOS imagery. Remote Sensing 2014, 6(12):12005-12036. https://doi.org/10.3390/rs61212005.

[7] Ibraheem NA, Khan RZ, and Hasan MM: Comparative study of skin color based segmentation techniques. International Journal of Applied Information Systems (IJAIS) 2013, 5(10).

[8] Cunningham SJ, Holmes G: Developing innovative applications in agriculture using data mining. 2001.

[9] Sathpathi S, Mohanty AK, Satpathi P, Mishra SK, Behera PK, and Patel G, Dondorp AM: Comparing Leishman and Giemsa staining for the assessment of peripheral blood smear preparations in a malaria-endemic region in India. Malaria journal 2014, 13(1):512. https://doi.org/10.1186/1475-2875-13-512.

[10] Prodanov D, Verstreken K: Automated segmentation and morphometry of cell and tissue structures. Selected algorithms in imageJ. In: Molecular Imaging. InTech; 2012. https://doi.org/10.5772/36729.

[11] Gautam A, Singh P, Raman B, Bhadauria H: Automatic classification of leukocytes using morphological features and naïve Bayes classifier. In: Region 10 Conference (TENCON), 2016 IEEE: 2016. IEEE: 1023-1027.

[12] Yang X, Shen X, Long J, Chen H: An improved median-based Otsu image thresholding algorithm. Aasri Procedia 2012, 3:468-473. https://doi.org/10.1016/j.aasri.2012.11.074.

[13] Gautam A, Bhadauria H: Classification of white blood cells based on morphological features. In: Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on: 2014. IEEE: 2363-2368.

[14] Gonzales R, Woods E: Digital Image Processing, 3, d edition. In.: Prentice-Hall; 2007.

[15] Barbedo JGA: Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus 2013, 2(1):660. https://doi.org/10.1186/2193-1801-2-660.

[16] Su M-C, Cheng C-Y, Wang P-C: A neural-network-based approach to white blood cell classification. The scientific world journal 2014, 2014.

[17] Raschka S: Naive bayes and text classification i-introduction and theory. arXiv preprint arXiv:14105329 2014.

[18] Amancio DR, Comin CH, Casanova D, Travieso G, Bruno OM, Rodrigues FA, da Fontoura Costa L: A systematic comparison of supervised classifiers. PloS one 2014, 9(4):e94137. https://doi.org/10.1371/journal.pone.0094137.

[19] Wu H, Phang TH, Liu B, Li X: A refinement approach to handling model misfit in text categorization. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining: 2002. ACM: 207-216. https://doi.org/10.1145/775047.775078.

[20] Kang H-W, Kang H-B: Prediction of crime occurrence from multimodal data using deep learning. PloS one 2017, 12(4):e0176244. https://doi.org/10.1371/journal.pone.0176244.