

Research Article

Unsupervised Leukocyte Image Segmentation Using Rough Fuzzy Clustering

Subrajeet Mohapatra, Dipti Patra, and Kundan Kumar

Department of Electrical Engineering, National Institute of Technology Rourkela, Orissa, Rourkela 769008, India

Correspondence should be addressed to Subrajeet Mohapatra, subrajeets@gmail.com

Received 6 October 2011; Accepted 21 November 2011

Academic Editor: C. Chen

Copyright © 2012 Subrajeet Mohapatra et al. 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.

The segmentation of leukocytes and their components acts as the foundation for all automated image-based hematological disease recognition systems. Perfection in image segmentation is a necessary condition for improving the diagnostic accuracy in automated cytology. Since the diagnostic information content of the segmented images is plentiful, suitable segmentation routines need to be developed for better disease recognition. Clustering is an essential image segmentation procedure which segments an image into desired regions. A judicious integration of rough sets and fuzzy sets is suitably employed towards leukocyte segmentation in a clustering framework. In this study, the goodness of fuzzy sets and rough sets is suitably integrated to achieve improved segmentation performance. The membership concept of fuzzy sets endow is efficient handling of overlapping partitions, and the rough sets provide a reasonable solution to deal with uncertainty, vagueness, and incompleteness in data. Such synergistic combination gives the proposed scheme an edge over standard cluster-based segmentation techniques, that is, K-means, K-medoid, fuzzy c-means, and rough c-means. Comparative analysis reveals that the hybrid rough fuzzy c-means algorithm is robust in segmenting stained blood microscopic images. The accomplished segmented nucleus and cytoplasm of a leukocyte can be used for feature extraction which leads to automated leukemia detection.

1. Introduction

Abnormal functioning of blood cells or blood-forming tissues is termed as hematological disorders. Cellular components of the blood are considered important, as the blood cells are easily accessible indicators of disturbances in their organs of origin or degradation which are much less accessible for diagnosis. Thus, changes in the erythrocyte, leukocytes, and platelets allow important inference to be drawn about various hematological disease conditions [1]. Visual (subjective) assessment of stained blood slides is a low cost, preferred, and reliable evaluation technique throughout the globe for initial screening of patients. Although significant improvements have been achieved in terms of identifying clinically relevant morphological clues, diagnostic hematology remains a subjective and time-consuming process. Human evaluation of blood slides is subjected to inter- and intraobserver variations resulting in poor diagnostic classification. It has been observed that cytological measurement

can significantly improve the diagnostic decision making. Such measurements are termed as quantitative microscopy and provides morphological changes in terms of numbers [2]. Quantitative measurement acts as an essential diagnostic tool for objective interpretation of hematological disorders like anemia, malaria, leukemia, AIDS, and so forth. The main objective of our studies is to deal with a specific neoplastic disorder of white blood cells (leukocytes) called leukemia. Leukemia is the neoplastic proliferations of hemopoietic cells and is considered as our subject of study. Leukemia can be pathologically understood as a hematological malignancy with increased numbers of myeloid or lymphoid blasts. Leukemia can be acute or chronic depending on the severity of the disease. Practical classification of leukemia is quite complicated and can be categorized on the basis of morphologic findings, genetic abnormalities, putative etiology, cell of origin, immunophenotypic qualities, and clinical characteristics. French, American, and British (FAB) classification and World Health Organization (WHO) classification are

two widely used protocols for leukemia categorization [3]. But both fundamentally divide leukemia's into myeloid and lymphoid types, depending on the origin of the blast cell. Acute lymphoblastic leukemia (ALL) is considered as the prime focus of our research.

ALL is the most common malignancy diagnosed in children representing nearly one third of all pediatric cancers [4]. Leukemia diagnosis serves as a pillar to all therapies, thus diagnostic tests are of utmost importance and need to be executed with precision. Single tests, that is, morphological, cytochemical, immunophenotyping, cytogenetic, and molecular genetic analysis or a combination of two tests is performed on the leukocytes for confirmation and classification of leukemia. As all other tests are expensive, microscopic examination of blood is an attractive diagnostic tool for initial screening of leukemia. Thus stained and fixed blood smears are extensively used for measuring and characterizing properties of the leukocytes based on shape variations of nucleus and cytoplasm for leukemia detection. Human evaluation is based on visual examination of the blood film based on their clinicopathological understanding and expertise [5]. Such techniques are prone to perverted results because of inter- and intraobserver variations and are also subjected to factors like slowness, operator tiredness, and so forth resulting-erroneous interpretation. In order to alleviate these bottlenecks, a computer-aided leukocyte segmentation mechanism is developed in this paper which facilitates automated leukemia detection.

Accurate segmentation of the cells is the first and necessary step for all automated cell analyzers. However, the complex biological structure of the leukocyte, poor staining, and touching cells makes the segmentation an ill-posed problem. Further the required extent of accuracy is very high in automated leukemia detection systems and solely depends on leukocyte segmentation. Segmentation of leukocytes is a complicated problem mostly because of unclear boundary between both cytoplasm and plasma (background) or cytoplasm and nucleus. Utmost care has to be taken while classifying the boundary pixels, as the roughness or irregular boundary is an essential feature for the diagnosis of leukemia.

Since there is no general solution to the image segmentation problem, specific algorithm has to be developed for segmenting leukocyte images. Standard segmentation procedures can be hybridized along with the domain knowledge to obtain desired segmentation results for specific problem domain. Promising segmentation results were obtained using fuzzy clustering technique for biological image samples [6–11]. The objective of blood image segmentation in the present context is to extract the morphological components such as nucleus and cytoplasm of each leukocyte using rough-fuzzy c-means (RFCM) clustering algorithm.

The rest of the paper is organized as follows. In the next section we briefly summarize the related works present in the literature. Section 3 describes the schema of the proposed method. Standard clustering techniques including partitive algorithms in the soft computing framework are outlined in Section 4. Section 5 provides a discussion on rough sets and its application to data clustering through

rough-fuzzy c-means (RFCM) algorithm. Proposed hybrid approach for leukocyte segmentation is outlined in Section 6. Experimental results are presented in Section 7. A detailed analysis of the results obtained is presented in Section 8. Concluding remarks are provided in Section 9.

2. Literature Survey

Over years numerous blood smear image segmentation methods have been proposed [12–14]. Those methods can be broadly classified as edge-based [15], region-based [16], threshold-based [17], and watershed-based [18, 19] segmentation schemes. Wu et al. [15] stated that cell boundaries are not sharp enough to perform edge-based segmentation in leukocyte images. An improved seeded region growing algorithm for cell segmentation was presented by Mehnert and Jackway [20]. However, determining the initial seed points is the drawback of all region-based methods. Few studies have been reported in the literature which employs thresholding for white blood cell (WBC) segmentation. Liao and Deng [21] introduced a gray level threshold based WBC image segmentation. Fuzzy divergence is employed by Ghosh et al. [5] for threshold estimation in leukocyte images. All threshold-based approaches are able to segment the nucleus from the background with acceptable accuracy. Color images are very rich source of information and regions can be segmented better in terms of color as compared to grayscale images. A two-step color image segmentation process using K-means clustering followed by EM algorithm was proposed by Sinha and Ramakrishnan [22]. Comaniciu and Meer [23] applied mean shift algorithm for color image segmentation of leukocyte images. Blood cell contour detection using active contour model was first presented by Kass et al. [24]. Another variation of active contour model (snakes) was explored successfully for WBC nucleus segmentation by Ongun et al. [25]. The application of morphological operators has also been investigated for WBC background separation [26]. In recent years, clustering technique has been incorporated intelligently for leukocyte image segmentation [6, 27]. As a general purpose segmentation method, feature space clustering has the advantage that is straight forward for classification [12]. Drawbacks associated with standard-clustering algorithms are the predetermination of number of clusters [28] and overlapping of morphological regions that is, cytoplasm and nucleus. As per Kumar et al. [29] selection of color space is also a vital issue in color-based clustering. Assumption of leukocytes as circular in shape-based methods is untenable in many cases; thus, diagnostic accuracy can drastically fall. There are several similar findings on blood-cell segmentation in the literature. It was also observed that standard segmentation methods are able to extract the WBC nucleus with acceptable level of accuracy but fails badly with cytoplasm. Cytoplasm is a decisive morphological component of blood; hence, utmost care should be taken while extracting it. To summarize the study reveals that the segmentation performances are limited by factors like smear preparation, staining, and image grabbing. So much work has to be done to meet real clinical demands.

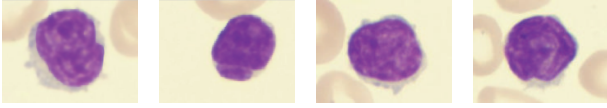


FIGURE 1: Sample-stained leukocytes.

Uncertainty arises due to color pixel similarity between cytoplasm and background (plasma) region. Misclassification of color pixel is an inherent problem in standard color-based clustering schemes [30]. In the present paper we devise a rough fuzzy set-based hybrid-clustering approach towards leukocyte segmentation in order to minimize these errors. Fuzzy *c*-means (FCM) [31] and rough *c*-means (RCM) [32] algorithms are merged together to develop a hybrid-clustering algorithm. Fuzzy sets have the ability to deal with issues like overlapping patterns, uncertainty, and vagueness. However, issues like incompleteness can be efficiently handled by rough sets. So rough-fuzzy *c*-means (RFCM) is an approach to merge the merits of FCM and RCM for large data clustering. The proposed scheme employs RFCM clustering to segment each leukocyte into its morphological components like cytoplasm and nucleus.

3. Material and Methods

3.1. Blood Smear Preparation. Blood samples were collected at Ispat General Hospital, Rourkela, India through randomization. Subsequently blood smear is prepared and stained using Leishman for visualization of cell components. The images were captured with a digital microscope (Carl Zeiss India) under 100x oil-immersed setting and with an effective magnification of 1000. Few images with permission from University of Virginia were also considered for experimental purposes. Figure 1 presents a set of sample-stained leukocyte images. The data set is a mixture of lymphocytes and lymphoblasts. There are 100 images collected from Ispat General Hospital, Rourkela, India, and 8 images are collected from University of Virginia. Manual segmentation was performed by Dr. Sanghamitra Satpathy, Hematologist, Department of Pathology, Ispat General Hospital, Rourkela, India. Each hand-segmented image consists of nucleus, cytoplasm, and back ground.

3.2. Subimaging. The input peripheral blood smear images are relatively larger with more than one leukocyte per image. As per the requirement, region of interest (ROI) must contain a single leukocyte only and is obtained by automatic cropping of the original input image. This is desired as every leukocyte in the input image has to be evaluated for classifying it as a blast cell. Thus, subimages containing single nucleus per image are obtained using bounding box [33] technique. We use simple K-means color-based clustering to obtain all the blue WBC nucleus of the entire image. Using image morphology we obtain the centroid of each nucleus, and a square image is cropped around each nucleus such that entire cell will be within the cropped subimage as shown in Figure 2. Again remapping with the original image, we can restore the color components and color subimages are

obtained and is shown in Figure 3. Subimages containing a single lymphocytes only were obtained and can now be used for further processing.

3.3. Preprocessing. Noise may be accumulated during image acquisition and due to excessive staining. All the test images are subjected to selective median filtering followed by unsharp masking [34]. Incorporation of adaptive threshold into the noise detection process led to more reliable and more efficient detection of noise. Minute edge details of the microscopic images are perfectly preserved even after median filtering. Unsharp masking is performed to sharpen the image details making the segmentation process easier.

3.4. Color Conversion. Typically images generated by digital microscopes are usually in RGB (Red, Green, and Blue) color space. A number of other color spaces or color models have been suggested in literature for various specific purposes. In the present paper we use $L^*a^*b^*$ color model for reduced color feature based clustering. The $L^*a^*b^*$ version of two sample images is shown in Figure 4. The $L^*a^*b^*$ color space is a color representation technique which consists of a luminosity layer L^* , chromaticity layer a^* , and chromaticity layer b^* . The color components, that is, a^* and b^* are used as features in the clustering process. Computation time is an important issue in all feature-based clustering problems with large data sets. Use of two color features (a^* and b^*) instead of three (red, green, and blue) reduces the computational time drastically.

3.5. Image Segmentation. Recognition of leukemia in blood samples is based on morphological variation of WBC. Such alterations can only be measured with segmented nuclei and cytoplasm. Accuracy of leukemia detection solely depends on leukocyte segmentation; thus, a suitable method has to be employed for morphological region extraction. Present paper deals with nucleus and cytoplasm region extraction from the background using rough-fuzzy *c*-means (RFCM) clustering. Detailed description of the proposed leukocyte image segmentation using rough-fuzzy *c*-means (RFCM) clustering is presented in Section 6. The obtained segmented regions can be used for feature extraction for acute leukemia detection. A brief overview of partitive clustering followed by an introduction to rough sets, rough *c*-means and rough-fuzzy *c*-means clustering, is presented in the following section.

4. Partitive Clustering

Clustering is an unsupervised classification of data patterns into homogeneous groups or clusters. It has been addressed by various researchers in diversified areas such as pattern recognition, data mining, image processing, biology, psychology, marketing, and so forth. This section provides an overview of widely used clustering techniques from an image segmentation perspective. Clustering techniques can be broadly categorized as

- (1) hard partitive clustering and
- (2) soft partitive clustering.

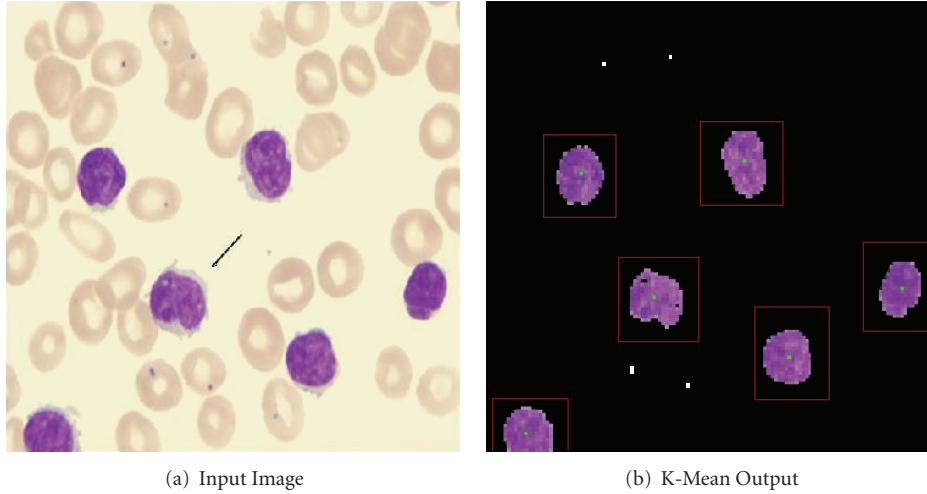


FIGURE 2: Initial K-means segmentation.

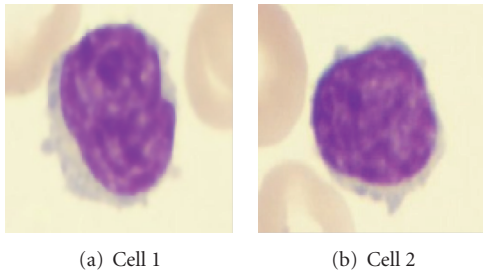
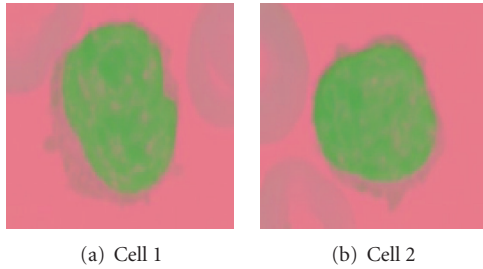


FIGURE 3: Cropped subimages.

FIGURE 4: Sub images in $L^*a^*b^*$ color space.

Popular clustering algorithms such as K-means, and K-medoid belong to the first category, where each data pattern is a member of exactly one cluster. Soft computing-based partitive clustering techniques broadly include fuzzy c-means (FCM), rough c-means (RCM). In this section introduction to standard-clustering algorithms is presented. Rough c-means (RCM) is presented along with an overview of rough sets in Section 5. Rough-fuzzy c-means (RFCM) clustering is presented in Section 5.2.

4.1. K-Means. K-means is a center-based clustering algorithm which is efficiently employed for clustering large databases and high-dimensional databases. The objective of a center-based algorithm is to minimize its objective function

is well suited for convex shape clusters, and fails drastically for clusters of arbitrary shapes [35]. The conventional K-means algorithm was first proposed by MacQueen (1967). This technique clusters the data into fixed number of clusters, and the mean of one cluster is placed as far away as possible from another. Every data point is associated to the nearest mean and belongs to one of the clusters [36]. Numerous variations of similar theme are also available in the literature which is usually based on changing the dissimilarity or centering.

4.2. K-Medoid. K-medoid is a similar clustering technique like K-means which tries to minimize a squared error criterion but the cluster center is chosen from the set of data points rather than mean. The element whose average dissimilarity to all the objects in the cluster is minimal is selected as medoid of that cluster [37]. It is immune to noise and outliers hence more suitable than K-means.

4.3. Fuzzy C-Means (FCM). The first algorithm in the soft partitive clustering arena was fuzzy c-means (FCM) and was developed in 1973 by Dunn and improved by [31]. In FCM each data point is associated with every cluster using a membership function, which gives degree of belongingness to the clusters. The partition matrix is obtained by minimizing an objective function

$$J = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^m \|X_k - v_i\|^2, \quad (1)$$

where $1 \leq m < \infty$ is the degree of fuzziness, v_i is the i th cluster center, $\mu \in [0, 1]$ is the membership of the k th data pattern to it, and $\|\cdot\|$ is the euclidean distance norm. Where as

$$v_i = \frac{\sum_{k=1}^N (\mu_{ik})^m X_k}{\sum_{k=1}^N (\mu_{ik})^m}, \quad (2)$$

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c (d_{ik}/d_{jk})^{2/(m-1)}}, \quad (3)$$

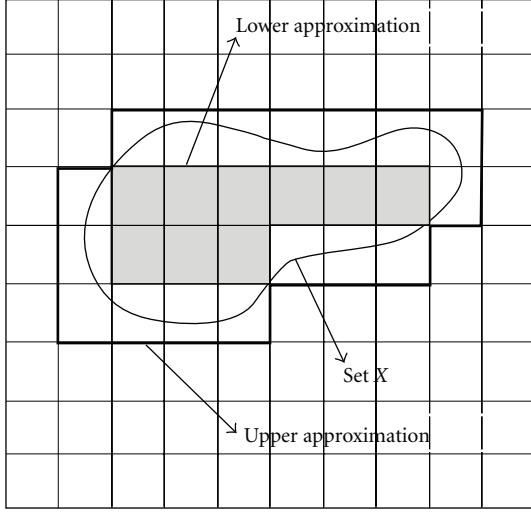


FIGURE 5: Lower and upper approximations in a rough set.

where $\forall i$ with $d_{ik} = \|X_k - v_i\|^2$, subject to $\sum_{i=1}^c \mu_{ik} = 1, \forall k$, and $0 < \sum_{i=1}^c \mu_{ik} < N, \forall i$. The FCM algorithm consists of the following steps.

- (1) Assign initial centroids $v_i, i = 1, 2, \dots, c$. Choose value of fuzzifier m and threshold t_{\max} . Set iteration counter $t = 1$.
- (2) Repeat steps 3-4 by incrementing t until $|\mu_{ik}(t) - \mu_{ik}(t-1)| > t_{\max}$.
- (3) Compute μ_{ik} by (3) for c clusters and N data patterns.
- (4) Update means, v_i , using (2).

Whereas Gustafson Kessel (GK) is a variation of FCM algorithm which associates each cluster with a cluster centre and with a covariance matrix. Original FCM implicitly considers each clustering data as spherical, while GK technique is not subjected to such assumptions and can also deal with nonspherical geometry of data.

5. Rough Sets

The principle of rough set is based on representation of rough or imprecise information in terms of exact concepts, that is, lower and upper approximation. These approximations (lower and upper) are obtained using an indiscernible relation based on the attributes of the objects in a domain. The set of objects which definitely belong to the vague concept are classified under lower approximation, whereas objects which possibly belong to the same are categorized as upper [38]. The difference of upper and lower approximation will result with objects in the rough boundaries. Figure 5 provides a schematic diagram of a rough set X within upper and lower approximation.

5.1. Rough C-Means (RCM). In Rough c-means (RCM) clustering, the idea of standard K-means is extended by visualizing each class as an interval or rough set [32]. A rough set Y

is characterized by its lower and upper approximations $\underline{B}Y$ and $\overline{B}Y$ respectively. In rough context an object X_k can be a member of at most one lower approximation. If $X_k \in \underline{B}Y$ of cluster Y , then concurrently $X_k \in \overline{B}Y$ of the same cluster. Whereas it will never belong to other clusters. If X_k is not a member of any lower approximation, then it will belong to two or more upper approximations. Updated centroid v_i of cluster U_i is computed as

$$v_i = \begin{cases} M_1, & \text{if } \underline{B}U_i \neq \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ M_2, & \text{if } \underline{B}U_i = \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ M_3, & \text{otherwise,} \end{cases} \quad (4)$$

where,

$$M_1 = w_{\text{low}} \frac{\sum_{X_k \in \underline{B}U_i} X_k}{|\underline{B}U_i|} + w_{\text{up}} \frac{\sum_{X_k \in (\overline{B}U_i - \underline{B}U_i)} X_k}{|\overline{B}U_i - \underline{B}U_i|},$$

$$M_2 = \frac{\sum_{X_k \in (\overline{B}U_i - \underline{B}U_i)} X_k}{|\overline{B}U_i - \underline{B}U_i|}, \quad (5)$$

$$M_3 = \frac{\sum_{X_k \in \underline{B}U_i} X_k}{|\underline{B}U_i|}.$$

The parameters w_{low} and w_{up} correspond to relative weighting factor for lower and upper approximation respectively towards centroid updation. In this process the weight factor for lower approximation ($\underline{B}U_i$) is higher than that of rough boundary ($\overline{B}U_i - \underline{B}U_i$), that is, $w_{\text{low}} > w_{\text{up}}$. Where $|\underline{B}U_i|$ signifies the number of members in the lower approximation of cluster U_i , where as $|\overline{B}U_i - \underline{B}U_i|$ is the number of members present in the rough boundary within the two approximations. The detailed RCM algorithm is presented below.

- (1) Assign initial centroids v_i for the c clusters.
- (2) Each data object X_k is assigned either to the lower approximation $\underline{B}U_i$ or upper approximation $\overline{B}U_i$ of cluster U_i , by computing the difference in its distance $d(X_k, v_i) - d(X_k, v_j)$ from cluster centroid pairs v_i and v_j .
- (3) **If** $d(X_k, v_i) - d(X_k, v_j)$ is less than a particular threshold T ,
then $X_k \in \overline{B}U_i$ and $X_k \in \underline{B}U_j$ and X_k cannot be a member of any lower approximation,
else $X_k \in \underline{B}U_i$ such that euclidean distance $d(X_k, v_i)$ is minimum over the c clusters.
- (4) Compute new updated centroid v_i for each cluster U_i using (4).
- (5) Iterate until convergence, that is, there are no more data members in the rough boundary.

Rough c-means algorithm is completely governed by three parameters such as w_{low} , w_{up} , and T . The parameter threshold can be defined as relative distance of a data member X_k from a pair of cluster centroids v_i and v_j . These parameters have to be suitably tuned for proper segmentation.

5.2. Rough-Fuzzy C-Means. Rough-fuzzy c -means [39] was developed by incorporating membership concept into RCM framework. In the present paper rough-fuzzy c -means algorithm is proposed for image segmentation. This permits for integrating fuzzy membership values μ_{ik} of a sample X_k to a cluster mean v_i , relative to all other means $v_j \forall j \neq i$, instead of absolute individual distance d_{ik} from the centroid as in RCM. Embedding fuzziness into RCM improves the robustness in clustering hence better segmentation accuracy can be achieved. The major steps of the algorithm is outlined below.

- (1) Assign initial centroids v_i for the c clusters.
- (2) Compute μ_{ik} using (3) for c clusters and N data objects.
- (3) Assign each data pattern X_k to the lower approximation $\underline{B}U_i$ or upper approximation $\overline{B}U_i, \overline{B}U_j$ of cluster pairs U_i and U_j by computing the difference in membership $\mu_{ik} - \mu_{jk}$.
- (4) Assuming μ_{ik} be maximum and μ_{jk} be the next to maximum.
If $\mu_{ik} - \mu_{jk}$ is less than some threshold,
then $X_k \in \overline{B}U_i$ and $X_k \in \underline{B}U_j$ and X_k cannot be a member of any lower approximation,
else $X_k \in \underline{B}U_i$ such that membership value μ_{ik} is maximum over the c clusters.
- (5) Compute updated centroid for each cluster U_i , incorporating (2) and (3) into (4), as in (6).
- (6) **Repeat** steps 2–5 **until** convergence, that is, there are no more new assignments

$$v_i = \begin{cases} M_1, & \text{if } \underline{B}U_i \neq \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ M_2, & \text{if } \underline{B}U_i = \emptyset \wedge \overline{B}U_i - \underline{B}U_i \neq \emptyset, \\ M_3, & \text{otherwise,} \end{cases} \quad (6)$$

where,

$$\begin{aligned} M_1 &= w_{\text{low}} \frac{\sum_{X_k \in \underline{B}U_i} \mu_{ik}^m X_k}{\sum_{X_k \in \underline{B}U_i} \mu_{ik}^m} + w_{\text{up}} \frac{\sum_{X_k \in (\overline{B}U_i - \underline{B}U_i)} \mu_{ik}^m X_k}{\sum_{X_k \in (\overline{B}U_i - \underline{B}U_i)} \mu_{ik}^m}, \\ M_2 &= \frac{\sum_{X_k \in (\overline{B}U_i - \underline{B}U_i)} \mu_{ik}^m X_k}{\sum_{X_k \in (\overline{B}U_i - \underline{B}U_i)} \mu_{ik}^m}, \\ M_3 &= \frac{\sum_{X_k \in \underline{B}U_i} \mu_{ik}^m X_k}{\sum_{X_k \in \underline{B}U_i} \mu_{ik}^m}. \end{aligned} \quad (7)$$

An optimal selection of above parameters is an important issue in rough-fuzzy c -means clustering. Similar to RCM, we use $w_{\text{up}} = 1 - w_{\text{low}}$, $0.5 < w_{\text{low}} < 1$, $0 < T < 0.5$ and $m = 2$.

6. Proposed RFCM Algorithm for Leukocyte Segmentation

Sub images containing a single leukocyte per image is desirable and is obtained as defined in Section 3.2. Blood images

generated from digital microscope are usually represented using RGB color model and contain three color bands, that is, red, green, and blue. Suitable color conversion from RGB to $L^*a^*b^*$ was done as defined in Section 3.4 to reduce the color dimension from three to two. Hence a^* and b^* component of the leukocyte image is considered as two feature inputs for color-based clustering. Leukocyte images can be visually segmented into four regions, that is, nucleus, cytoplasm, red blood cells (RBCs), and background stain as suggested by the hematologist. Inconsistency in color variation within the nucleus is also an issue which increases the total number of visual classes to five. Experiments were conducted to determine the exact number of classes c for accurate segmentation of the leukocyte. After rigorous empirical study, the number of classes c was found to be four. Due to unequal color variation of the stain within the nucleus is represented as two separate regions. Cytoplasm and background stain which also include RBC are considered as other two regions.

Rough-fuzzy c -means (RFCM) clustering is employed to classify each pixel into four clusters. The proposed segmentation algorithm is applied on each subimage to separate the nucleus and cytoplasm from the background. The detailed algorithm is as follows.

- (1) Let I_{rgb} represent an original color leukocyte image in RGB color format.
- (2) Apply $L^*a^*b^*$ color space conversion on I_{rgb} to obtain the $L^*a^*b^*$ image that is, I_{lab} .
- (3) Construct the input feature vector using a^* and b^* components of I_{lab} .
- (4) Each data pattern of the feature vector is assigned to a appropriate class using rough-fuzzy c -means algorithm.
- (5) Obtain the labeled image from the classified feature vector.
- (6) Reconstruct the segmented RGB color image for each class.

After segmentation each pixel of the leukocyte image is classified as one of the four clusters based on corresponding a^* and b^* values in $L^*a^*b^*$ color space. Clustered output in terms of a scatter plot for the image (Figure 4(a)) is shown in Figure 6.

7. Simulation Results

The efficacy of the proposed scheme is demonstrated by conducting four experiments on the entire set of available images which is 100 for our case. However due to space constraint experimental results for two lymphocyte images only are presented in the current section. Segmentation performance in terms of visual assessment is demonstrated through the first experiment. Clustering performance in terms of cluster validity index that is, global silhouette index (SL) [40] and partition index (SC) [41] is presented through the second experiment for establishing quantitative performance evidence. The third experiment deals with the

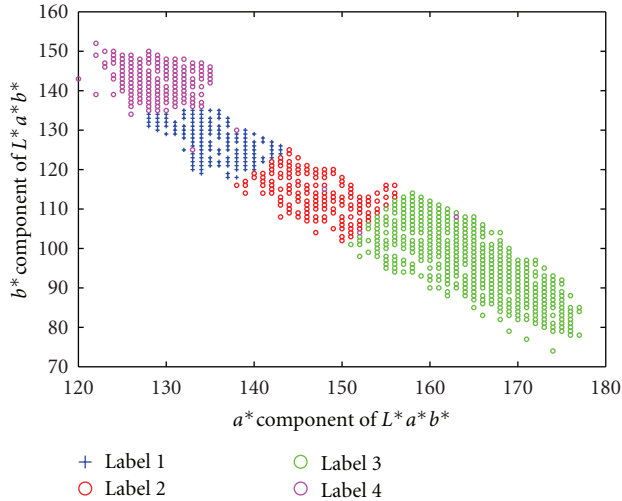


FIGURE 6: Feature space clustering result for rough fuzzy C means.

comparative analysis of the proposed technique in terms of misclassification error with the available hand segmented images. In the last experiment, proposed scheme is compared with the reported schemes in terms of computation time.

7.1. Experiment 1. Leukocyte image samples of size 128×128 as shown in Figures 3(a) and 3(b) are mapped from RGB color space to $L^*a^*b^*$ color space as shown in Figures 4(a) and 4(b). The color information in the $L^*a^*b^*$ color space is represented using two components (a^* and b^*) only. This property of reduction in number of color features from three to two can be utilized in accelerating color-based clustering process. Thus a^* and b^* component for every pixel is recorded, and feature data set X of size 16384×2 is prepared. Each row of X represents a data pattern, and redundancy among them was discarded. This concise form of X with size $N \times 2$ serves as an input towards pixel-labeling problem through color-based clustering.

After successful clustering, background including RBC is clustered into single class whereas cytoplasm is considered in another class. However, the entire nucleus is represented in two different clusters due to inconsistency in absorption of the staining material. Various standard clustering schemes such as K-means, K-medoid, Fuzzy c -means (FCM), Gustafson Kessel (GK), Rough c -means (RCM) are simulated along with our proposed scheme for obtaining the corresponding individual clusters. Segmented results obtained from different clustering schemes are presented in Figure 7 for the first leukocyte image sample (Figure 3(a)) and in Figure 8 for the second image sample (Figure 3(b)). Each column represents a particular cluster, and each row of the image indicates a particular clustering scheme. As we have four clusters, the image indicates four cluster outputs for each clustering scheme.

7.2. Experiment 2. Clustering algorithms are very sensitive to the type of data set and especially to noise and dimension. Cluster validity indexes have been used to evaluate the fitness

TABLE 1: Clustering performance for cell 1.

Clustering method		SL	SC
Hard Computing	K-means	0.1656	0.0308
	K-medoid	0.15232	0.2741
Fuzzy computing	FCM	0.15723	0.2219
	GK	0.1848	0.2298
	RCM	0.1970	0.0114
	Proposed	0.3624	0.0018

TABLE 2: Clustering performance for cell 2.

Clustering method		SL	SC
Hard Computing	K-means	0.1672	0.0501
	K-medoid	0.1536	0.2069
Fuzzy Computing	FCM	0.1546	0.2094
	GK	0.1650	0.1535
	RCM	0.1984	0.0102
	Proposed	0.3014	0.0026

of partitions obtained by various clustering algorithms [42]. Standard cluster validity index that is, global silhouette index, and partition index are used as a performance measure. The input $L^*a^*b^*$ images as shown in Figures 4(a) and 4(b) are segmented using all the clustering techniques independently as discussed in Section 4. Corresponding global silhouette index (SL) and partition index (SC) for each clustering technique was measured for optimum number of clusters as four and is tabulated in Tables 1 and 2, respectively.

7.3. Experiment 3. In this experiment all the standard color-based clustering schemes are applied to both the sample images, and segmentation performance is measured in terms of misclassification error. Since the predefined regions of the ground truth image (Figures 9(a) and 9(b)) is available from the hematologist the error rate can be computed for each region (cytoplasm, nucleus, and background) separately using the relation,

$$\varepsilon = \frac{\text{Total number of misclassified pixels}}{\text{Total number of pixels in a region}}, \quad (8)$$

where ε is the error rate. Individual clusters representing nucleus were added to obtain the desired nucleus image, and the misclassification error for the nucleus region along with the other regions was determined (see Tables 3 and 4).

7.4. Experiment 4. Both the sample leukocyte images are subjected to segmentation with all existing schemes along with the proposed schemes. The computational time (in seconds) are recorded for all the schemes and shown in Figure 10. It is observed that the proposed RFCM technique is computationally slower than standard K-means, K-medoids, FCM and GK algorithms and faster than RCM. However, the segmentation performance is much superior to those standard schemes.

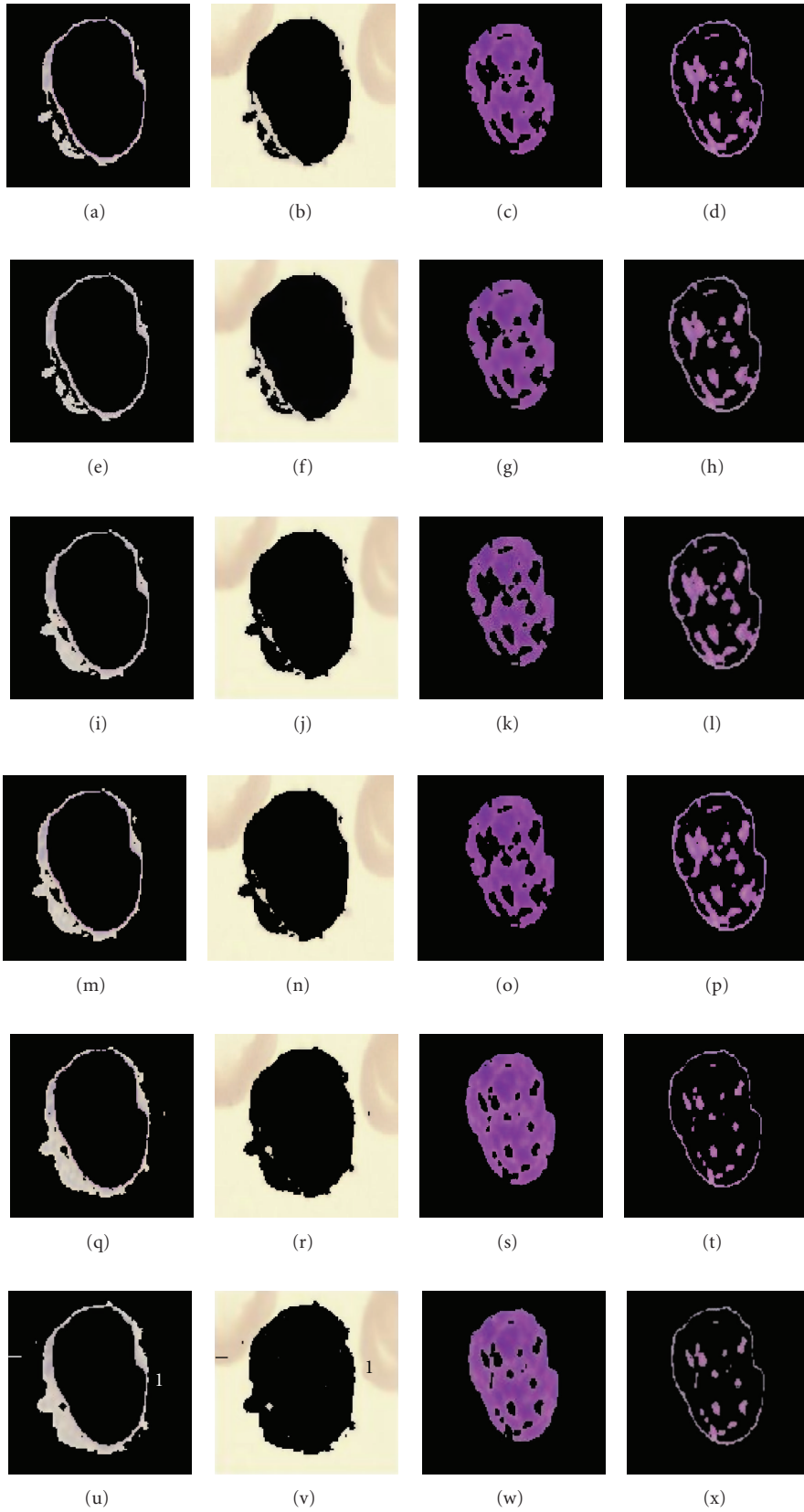


FIGURE 7: Clustering results formed by different clustering techniques. K-Means (a–d), K-Medoid (e–h), FCM (i–l), GK (m–p), RCM (q–t), Proposed (u–x).

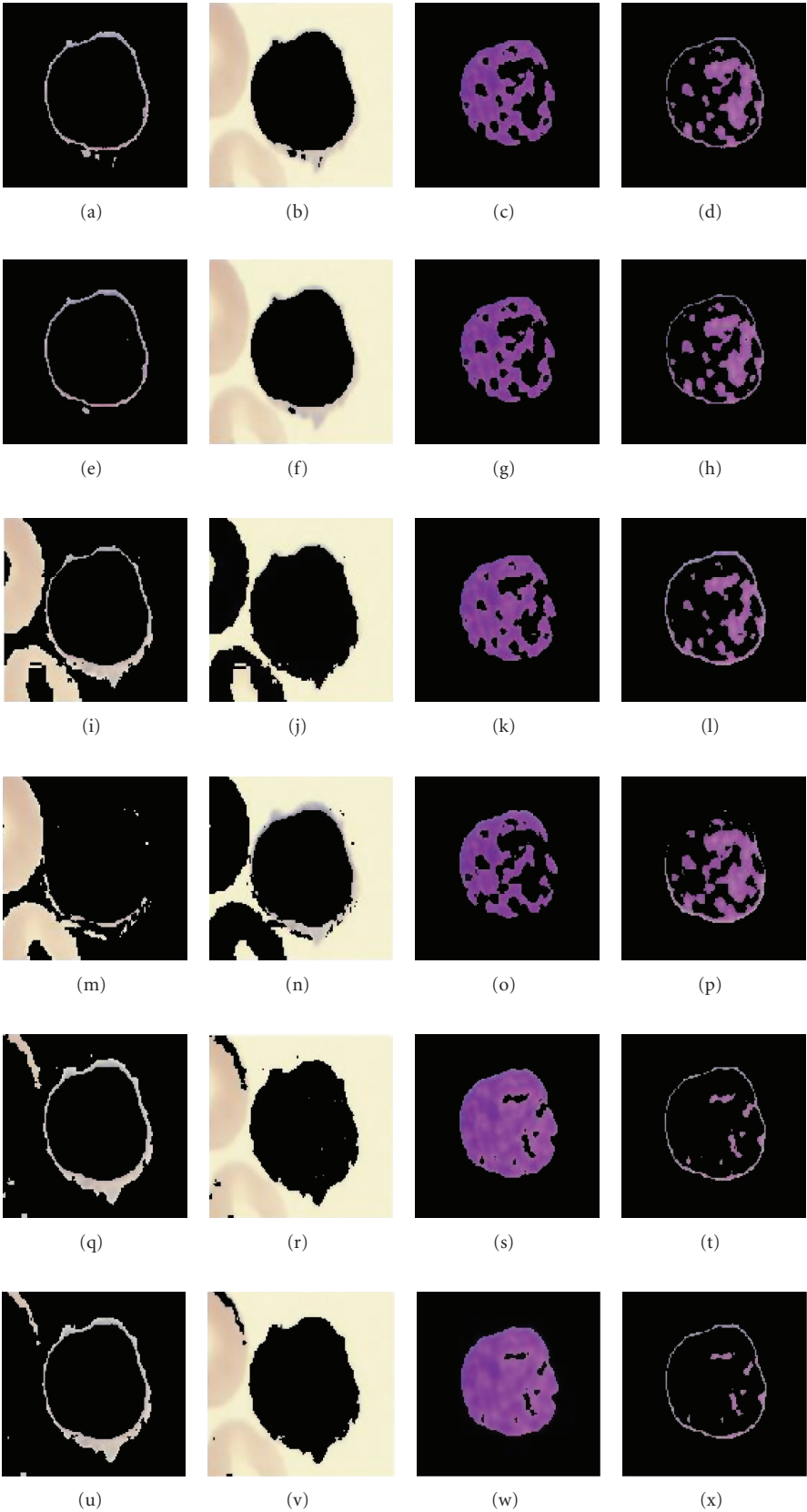


FIGURE 8: Clustering results formed by different clustering techniques. K-means (a-d), K-medoid (e-h), FCM (i-l), GK (m-p), RCM (q-t), proposed (u-x) for blood cell image sample 2.

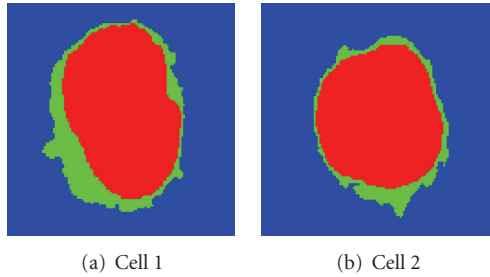


FIGURE 9: Ground truth image of the leukocyte samples.

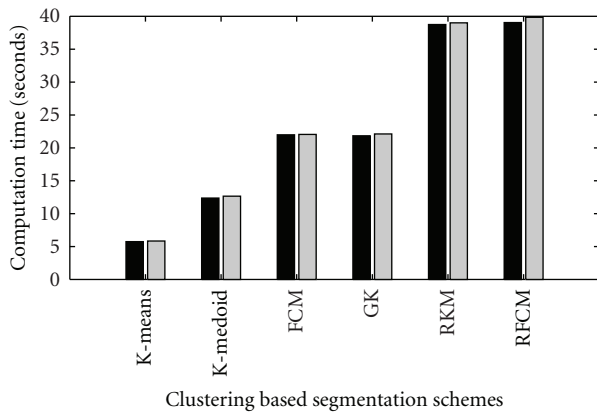


FIGURE 10: Variation of computational time in seconds.

TABLE 3: Misclassification error percentage for cell 1.

Schemes	Cytoplasm	Nucleus	Background
K-means	37.26	0.19	5.84
K-medoid	34.53	0.43	5.53
FCM	16.98	0.22	2.52
GK	13.18	2.17	1.05
RCM	7.98	0.87	0.42
Proposed	3.67	1.28	1.14

TABLE 4: Misclassification error percentage for cell 2.

Schemes	Cytoplasm	Nucleus	Background
K-means	45.49	5.08	5.86
K-medoid	50.73	2.53	5.32
FCM	303.98	1.44	26.62
GK	242.66	2.97	21.95
RCM	12.36	1.41	1.77
Proposed	10.27	1.32	1.39

8. Analysis

Automatic leukemia detection from leukocyte images is only possible by morphological analysis of nucleus and cytoplasm region individually. Accuracy of detection solely depends on nucleus and cytoplasm region extraction from the leukocyte image. Utilization of a suitable segmentation

technique drastically improves the diagnosis accuracy and is very essential for any medical image analysis system. Segmenting nucleus and cytoplasm is a very difficult task, and most of the reported schemes are able to extract the nucleus only. Cytoplasm is also an essential indicator of disease condition which has to be extracted for automatic disease recognition. Thus, RFCM clustering was employed for accurate leukocyte image segmentation and to extract the nucleus and cytoplasmic region under the clustering framework. The proposed scheme is computationally slower; however, the clustering performance in terms of cluster validity index (PC and SC) was found to be superior in comparison to the existing schemes. Further the proposed approach outperforms the other reported schemes in terms of misclassification error rate. Due to unavailability of standard segmentation performance measure, visual assessment was performed for the proposed scheme and was found to be outstanding in terms of cytoplasm extraction. Experimental results reveal that the proposed scheme outperforms all other reported schemes in terms of cytoplasm extraction along with satisfactory nucleus separation. Further this technique is computationally equivalent in comparison to RCM approach with significant segmentation performance. Thus such a hybrid approach towards leukocyte segmentation will facilitate accurate leukemia recognition. Similar test was performed on the entire available data set of 108 images, and satisfactory results were obtained.

9. Conclusion and Future Work

This paper proposes a rough-fuzzy hybrid-clustering technique for leukocyte image segmentation. The goodness of rough sets and fuzzy sets were suitably incorporated in the clustering framework to provide better segmentation performance. Encouraging segmentation results were obtained for images collected from two different locations. Exhaustive simulation on different cell images is performed, and it was observed that the cytoplasm and nucleus regions can be very well extracted using the proposed technique. Both subjective and objective comparative analysis with the existing standard schemes reveals that the proposed scheme outperforms others. Results obtained stimulate future works which includes reducing computational time and segmentation of blood smear images for overlapping leukocytes.

References

- [1] H. Thieml, H. Diem, and T. Haferlach, *Color Atlas of Hematology*, Thieme, 2004.
- [2] D. Burnett and J. Crocker, *The Science of Laboratory Diagnosis*, John Wiley & Sons, New York, NY, USA, 2005.
- [3] C. D. Tkachuk and J. V. Hirschmann, *Wintrobe's Atlas of Clinical Hematology*, Lippincott Williams & Wilkins, Philadelphia, Pa, USA, 1st edition, 2007.
- [4] N. Satake and J. M. Yoon, "Acute lymphoblastic leukemia," 2010, <http://emedicine.medscape.com/>.
- [5] M. Ghosh, D. Das, C. Chakraborty, and A. K. Ray, "Automated leukocyte recognition using fuzzy divergence," *Micron*, vol. 41, no. 7, pp. 840–846, 2010.

- [6] N. T. Umpon, "Patch based white blood cell nucleus segmentation using fuzzy clustering," *ECTI Transaction Electrical Electronics Communications*, vol. 3, no. 1, pp. 5–10, 2005.
- [7] W. Shitong and W. Min, "A new detection algorithm (NDA) based on fuzzy cellular neural networks for white blood cell detection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 5–10, 2006.
- [8] E. Montseny, P. Sobrevilla, and S. Romani, "A fuzzy approach to white blood cells segmentation in color bone marrow images," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 173–178, 2004.
- [9] N. Theera-Umpon and S. Dhompongsa, "Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 353–359, 2007.
- [10] A. I. Shihab, *Fuzzy clustering algorithms and their application to medical image analysis*, Ph.D. thesis, University of London, 2000.
- [11] K. I. Laws, *Texture image segmentation*, Ph.D. thesis, University of South California, 1980.
- [12] R. Adollah, M. Mashor, N. M. Nasir, H. Rosline, H. Mahsin, and H. Adilah, "Blood cell image segmentation: a review," in *Proceedings of the 4th Kuala Lumpur International Conference on Biomedical Engineering*, N. A. Osman, F. Ibrahim, W. W. Abas, H. A. RahmanTing, and H. Ting, Eds., vol. 21, pp. 141–144, Springer, 2008.
- [13] C. Di Rubeto, A. Dempster, S. Khan, and B. Jarra, "Segmentation of blood images using morphological operators," in *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 3, pp. 397–400, 2000.
- [14] D. Anoraganingrum, "Cell segmentation with median filter and mathematical morphology operation," in *Proceedings of the International Conference on Image Analysis and Processing*, pp. 1043–1046, 1999.
- [15] J. Wu, P. Zeng, Y. Zhou, and C. Olivier, "A novel color image segmentation method and its application to white blood cell image analysis," in *Proceedings of the 8th International Conference on Signal Processing*, vol. 2, 2006.
- [16] T. Mouroutis, S. J. Roberts, and A. A. Bharath, "Robust cell nuclei segmentation using statistical modelling," *Bioimaging*, vol. 6, no. 2, pp. 79–91, 1998.
- [17] H. S. Wu, J. Barba, and J. Gil, "Iterative thresholding for segmentation of cells from noisy images," *Journal of Microscopy*, vol. 197, no. 3, pp. 296–304, 2000.
- [18] G. Lin, U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roysam, "A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks," *Cytometry Part A*, vol. 56, no. 1, pp. 23–36, 2003.
- [19] M. Ghosh, D. Das, S. Mandal et al., "Statistical pattern analysis of white blood cell nuclei morphometry," in *Proceedings of the IEEE Students' Technology Symposium (TechSym '10)*, pp. 59–66, April 2010.
- [20] A. Mehnert and P. Jackway, "An improved seeded region growing algorithm," *Pattern Recognition Letters*, vol. 18, no. 10, pp. 1065–1071, 1997.
- [21] Q. Liao and Y. Deng, "An accurate segmentation method for white blood cell images," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 245–258, 2002.
- [22] A. Sinha and A. Ramakrishnan, "Automation of differential blood count," in *Proceedings of the Conference on Convergent Technologies for Asia-Pacific Region*, pp. 547–551, 547–551, 2003.
- [23] D. Comaniciu and P. Meer, *Cell Image Segmentation for Diagnostic Pathology*, Springer, New York, NY, USA, 2001.
- [24] M. Kass, A. Witkins, and D. Terzopoulos, "Snakes: active contour models," in *Proceedings of the 1st International Conference on Computer Vision*, pp. 259–268, 1987.
- [25] G. Ongun, U. Halici, K. Leblebicioğlu, V. Atalay, S. Beksac, and M. Beksac, "Automated contour detection in blood cell images by an efficient snake algorithm," *Nonlinear Analysis, Theory, Methods and Applications*, vol. 47, no. 9, pp. 5839–5847, 2001.
- [26] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," *Image and Vision Computing*, vol. 20, no. 2, pp. 133–146, 2002.
- [27] G. Ongun, U. Halici, K. Leblebicioğlu, V. Atalay, M. Beksac, and S. Beksac, "A modified fuzzy clustering for white blood cell segmentation," in *Proceedings of the 3rd International Symposium on Biomedical Engineering*, pp. 356–359, 2008.
- [28] K. Jiang, Qing-Min, and S.-Y. Dai, "Red blood cell segmentation scheme utilizing various image segmentation techniques," in *Proceedings of the 2nd International Conference on Machine Learning and Cybernetics*, 2003.
- [29] B. R. Kumar, D. K. Joseph, and T. Sreenivas, "Teager energy based blood cell segmentation," in *Proceedings of the 14th International Conference on Digital Signal Processing*, Bangalore, India, 2002.
- [30] S. Mohapatra and D. Patra, "Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images," in *Proceedings of the International Conference on Systems in Medicine and Biology (ICSMB '10)*, pp. 49–54, 2010.
- [31] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, USA, 1981.
- [32] P. Lingras and C. West, "Interval set clustering of web users with rough K-means," *Journal of Intelligent Information Systems*, vol. 23, no. 1, pp. 5–16, 2004.
- [33] A. K. Jain, *Fundamentals of Digital Image Processing*, Pearson Education, India, 1st edition, 2003.
- [34] S. Mohapatra, *Deveopmant of impulse noise detection schemes for selective filtering*, M.S. thesis, National Institute of Technology Rourkela, 2008.
- [35] G. Gan, C. Ma, and J. Wu, *Data Clustering Theory, Algorithms, and Applications*, Society for Industrial and Applied Mathematics, 2007.
- [36] N. K. Verma, A. Roy, and S. Vasikarla, "Medical image segmentation using improved mountain clustering technique version-2," in *Proceedings of the IEEE 7th International Conference on Information Technology*, pp. 156–161, 2010.
- [37] B. Clarke, E. Fokue, and H. H. Zhang, *Principles and Theory for Data Mining and Machine Learning*, Springer, New York, NY, USA, 2009.
- [38] S. Mitra, "An evolutionary rough partitive clustering," *Pattern Recognition Letters*, vol. 25, no. 12, pp. 1439–1449, 2004.
- [39] S. Mitra, H. Banka, and W. Pedrycz, "Rough-fuzzy collaborative clustering," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 4, pp. 795–805, 2006.
- [40] B. Risteovski, S. Loshkovska, S. Dzeroski, and I. Slavkov, "A comparison of validation indices for evaluation of clustering results of dna microarray data," in *Proceedings of the The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE '08)*, pp. 587–591, 2008.

- [41] K. L. Wu, "An analysis of robustness of partition coefficient index," in *Proceedings of the IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence) (FUZZ-IEEE '08)*, pp. 372–376, 2008.
- [42] K.-L. Wu and M.-S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275–1291, 2005.




Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

