

A Comparative Analysis of Depth Computation of Leukaemia Images using a Refined Bit Plane and Uncertainty Based Clustering Techniques

Swarnalatha Purushotham, B. K. Tripathy

School of Computing Science and Engineering, VIT University, Vellore – 632014, India

Emails: pswarnalatha@vit.ac.in tripathybk@vit.ac.in

Abstract: *Several image segmentation techniques have been developed over the years to analyze the characteristics of images. Among these, the uncertainty based approaches and their hybrids have been found to be more efficient than the conventional and individual ones. Very recently, a hybrid clustering algorithm, called Rough Intuitionistic Fuzzy C-Means (RIFCM) was proposed by the authors and proved to be more efficient than the conventional and other algorithms applied in this direction, using various datasets. Besides, in order to remove noise from the images, a Refined Bit Plane (RBP) algorithm was introduced by us. In this paper we use a combination of the RBP and RIFCM to propose an approach and apply it to leukemia images. The aim of the paper is twofold. First, it establishes the superiority of our approach in medical diagnosis in comparison to most of the conventional, as well as uncertainty based approaches. The other objective is to provide a computer-aided diagnosis system that will assist the doctors in evaluating medical images in general, and also in easy and better assessment of the disease in leukaemia patients. We have applied several measures like DB-index, D-index, RMSE, PSNR, time estimation in depth computation and histogram analysis to support our conclusions.*

Keywords: *RBP, Otsu thresholding, histogram analysis, RMSE, PSNR, clustering, conventional methods, C-Means, RIFCM, depth computation, leukaemia images.*

1. Introduction

Leukaemia is a cancerous condition characterized by a large quantity of abnormal white blood cells in the body. It affects blood forming cells in the body. Leukaemia initiates in the bone marrow and spreads to other parts of the patient's body.

There are two types of leukemia, namely acute and chronic. In chronic leukaemia, the leukaemia cells multiply slowly coming from mature, abnormal cells. Acute leukaemia develops from early cells, called "blasts", which are young cells that get divided frequently. The causes of leukemia can arise abruptly or slowly and the symptoms are fever, weakness, infections, etc.

Doctors can diagnose leukaemia from the results of several tests ranging from blood tests to spinal taps. A sample of blood is examined to check for changes in the chromosomes of the lymphocytes. Several tests are conducted concerning the functioning of different organs, in order to confirm whether a person is affected by leukaemia or not.

Treatment for leukaemia varies greatly depending on its type and the stage of the disease. Many times leukaemia is treated by one or more types of treatments. These are chemotherapy, radiation therapy, biologic therapy, surgery and hematopoietic cells or bone marrow transplant, based on the seriousness of the disease. Except hematopoietic transplant which is a procedure to restore the normal marrow production that has been destroyed by treatment with high doses of anticancer drugs or radiation, others do not need any method of replacement. Hence, classification plays a major role in the diagnosis of the disease and the treatment is based on its degree of development.

Medical imaging is becoming widely prevalent for diagnosis and analysis of a varied number of diseases. Image processing techniques are now being extended for early and efficient detection of blood cancer or leukaemia.

As above noted, leukaemia affects the blood forming cells and it results in the production of abnormal leucocytes or White Blood Cells (WBC) in a large amount. Since the number of these cells increases, they hinder the production of the Red Blood Cells (RBC) and proliferate to the bone marrow [1]. When the number of abnormal blood cells increases in the bone marrow, they overflow and enter the blood vessels, leading to the transfer of these cells to other organs and causing multi-organ failure. The presence of a large number of white blood cells indicates the presence of leukaemia.

Conventional methods use manual counting of the blood cells from the blood smear of the patient. This method is very slow. Moreover, the technicians need to start all over again if they are interrupted inbetween. To overcome this problem, we need a feasible solution which is not only faster than the traditional methods, but it is also accurate and cost-efficient.

In this paper we shall use an algorithm called a RBP filter algorithm for the removal of distortions from leukaemia images and enhancing them. This algorithm divides the image in a set of 8 bit planes, each one corresponding to a bit position ranging from 0 to 7. This sliced image is used for enhancement.

The edge detection methods can be used for segmenting leukaemia images into edges. There are four different conventional methods – Canny (C), Sobel (S), Zero-Cross (Z) and Robert-Cross (R). False edge detection generates thin or thick lines indicating the presence of noise. The Refined Bit Plane (RBP) algorithm filters out useless information, by preserving the significant structural properties in a leukaemia image. These enhanced leukaemia images are helpful for applications in depth computation [2], image reconstruction [3], etc.

Depth computation is carried out by finding the third dimension of a clustered Leukaemia image using conventional methods. The clustered Leukaemia image is necessary for further processing during image reconstruction. The damaged portions can be identified by finding the intensity values in the region of extraction and comparing them with the threshold value. If the intensity value is less than the threshold value then it is considered damaged [4, 5].

Since the conventional methods are not suitable to handle data having inherent uncertainty, FCM, IFCM, RCM and their hybrid means (Rough Fuzzy C-Means) RFCM and RIFCM clustering techniques are used in this paper. To be more precise, in this paper we use nine clustering techniques (both conventional and uncertainty based) along with the RBP algorithm for better segmentation of leukaemia images. We also carry out the segmentation process without RBP. Compared to all the other eight methods, the Rough Intuitionistic Fuzzy C-Means with a Refined Bit Plane (RIFCMRBP) yields better segmented images. Depth computation is essential to determine the damaged portions of a leukaemia image. The original image on depth computation takes more time than the images enhanced by using RBP. Thus, the time complexity for diagnosing leukaemia images using RIFCMRBP method is less.

The skeleton of the paper is as follows. References survey of the problem under study is carried out in Section 2. This discussion also covers brief introductions to RBP method, the conventional and uncertainty based clustering techniques. Section 3 deals with definitions of the concepts used in the paper. In Section 4 we describe our proposed methodology, conventional techniques, a refined bit plane filter and the various C-Means techniques, including RIFCM. We also introduce the DB and Dunn accuracy measures. In Section 5 we present a comparative analysis of all the nine clustering techniques based on the experimental results, taking standard input images. In Section 6 we deal with depth computation which leads to the measure of time complexity. In Section 7 we provide concluding observations and also present some possible future enhancements. We compile the details of the source materials consulted during the preparation of this work as references.

2. References survey

In human blood, the leukocytes consist of lymphocytes, monocytes, eosinophils, basophiles and neutrophils. These are differentiated by their cytoplasmic granules, staining properties of the granules, size of the cell, the proportion of the nuclear to the cytoplasmic material and the type of nucleolar lobes. There may be an increase

or decrease in the number of leukocytes due to infection or acute stress. In [6] an automatic recognition system of leukocytes which will extract the leukocyte region from a blood smear image which reduces the generation of leukocytes, has been developed.

A cluster is a collection of data elements that are similar to each other but dissimilar to elements in other clusters. A wide number of clustering algorithms have been proposed to suit the requirements in each application field. Cluster analysis is a popular data mining method being applied to fields like web mining [7], biology ([8, 9]), image processing [10] and market segmentation [11]. Clustering algorithms help to discover the intrinsic structural complexity of a dataset.

Image segmentation is carried out by using clustering techniques. In image segmentation, the basic features which can be segmented are image luminance amplitude for a monochrome image, color components for a color image and edges and texture of the image. Medical images are used for study of the anatomical structure, diagnosis, treatment planning or locating tumors, etc. In [12] a comparative study is made for acute leukaemia image segmentation using HIS and RGB colour space.

In [13] a novel Shadowed C-Means (SCM) clustering approach is applied on segmenting leukaemia cells in blood microscopic images and the proposed algorithm has been established to be fast and robust in segmenting stained blood microscopic images in the presence of outliers.

To improve the accuracy of segmentation, an Interest-based Ordering Scheme (IOS) for fuzzy morphology on White-Blood-Cells (WBC) has been proposed in [14]. In [15], using a refined bit plane method, a lossless hybrid algorithm based on a simple choosy scan order is presented for medical images, leukaemia images, etc.

Hard C-means algorithm [16], which is the first among the data clustering algorithms, assigns each object to one cluster. It exhibits the advantage of fast convergence rate, but still has some difficulties in dealing with overlapping or skewed data distributions problems [17]. However, the uncertainty nature of real-world datasets has necessitated the development of algorithms generating overlapping clusters. Perhaps the earliest algorithm having this characteristic is the fuzzy C-means algorithm developed by Bezdek in 1981 [18]. In this case an object can be assigned into many clusters with different membership values. Rough set theory provides another approach to deal with uncertainty and the first algorithm using rough sets for data clustering, called rough C-means was developed by Lingras and West [19]. Based on the rough set theory [20], a cluster is strictly defined by means of two approximations, called lower approximation and upper approximation of a cluster.

Rough set theory, which was supposed to be a rival for fuzzy set theory at the beginning, was shown to be actually a complement to it by Dubois and Prade [21]. In fact, they developed the hybrid models of rough fuzzy sets and fuzzy rough sets. Later on, an integrated technology called Rough Fuzzy C-Means (RFCM) algorithm was proposed in [22]. An application of this method was developed to analyze MRI medical images [23]. Although rough set-based clustering algorithms

have been considered efficient soft computing methods, a prevalent problem that appears is the parameter selection. Since no prior knowledge is given, an appropriate parameter may not be easily chosen. No matter what the actual distribution of each cluster is, a constant weighted parameter is always set manually.

The notion of intuitionistic fuzzy sets, introduced by Atanassov [24] generalizes the notion of fuzzy sets. Using it, the IFCM algorithm was introduced by Chaira and Chen [25]. Recently, Tripathy et al. [26] have introduced the RIFCM algorithm and studied their properties through some experimental analysis. It is established, using some accuracy measuring indices, like the Dunn index and DB index, that RIFCM outperforms all the existing algorithms in this aspect.

In this paper we use conventional methods, as well as imprecision based models like FCM, IFCM, RCM and their hybrid models like RFCM and RIFCM to the original leukaemia images, as well as their filtered versions obtained by using the algorithm called the RBP algorithm. It is established that the performance of RIFCMRBP is better than other techniques using several techniques including efficiency measuring indices like DB-index and D-index. Besides, it has the least time complexity among all the algorithms considered. It may be noted that another application of RIFCMRBP for analysis of satellite images has been discussed in [27].

3. Definitions of concepts

In this section we introduce some definitions to be used throughout the paper. In this paper we shall use the models of a Fuzzy Set (FS), an Intuitionistic Fuzzy Set (IFS), a Rough Set (RS), a Rough Fuzzy Set (RFS) and a Rough Intuitionistic Fuzzy Set (RIFS). We introduce these concepts below. Firstly, we define the notion of a fuzzy set, introduced by Zadeh [28].

Definition 3.1. A fuzzy set A defined over a universal set U is characterised by its membership function $\mu_A:U \rightarrow [0, 1]$, such that each $x \in U$ is associated with its membership value $\mu_A(x)$, which lies within the unit interval $[0, 1]$.

The notion of a fuzzy set was extended to define the concept of an intuitionistic fuzzy set by Atanassov [24], as follows.

Definition 3.2. An intuitionistic fuzzy set A defined over a universal set U is characterised by two functions, called membership and non-membership functions, given by $\mu_A, \nu_A:U \rightarrow [0, 1]$, such that each $x \in U$ is associated with two real numbers $\mu_A(x)$ and $\nu_A(x)$ lying in the unit interval $[0, 1]$ such that $0 < \mu_A(x) + \nu_A(x) \leq 1$.

It is worth noting that every intuitionistic fuzzy set A is associated with a function π_A defined through membership and non-membership functions, such that $\forall x \in U, \pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$.

The notion of a rough set was introduced by Pawlak [20] as another model to handle uncertainty in data. It is defined as follows.

Let U be a universe and P be a family of equivalence relations defined over U . The tuple (U, P) is called a knowledge base. Each $R \in P$ denotes the set of equivalence classes generated by R by U/R . For any $K \subseteq P$, $\text{IND}(K)$ denotes the intersection of all the equivalence relations in K and it is itself an equivalence relation over U . We define $\text{IND}(P) = \{\text{IND}(K) : K \subseteq P\}$.

Definition 3.3. For any $X \subseteq U$ and $R \in \text{IND}(P)$ we define two sets $\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$ and $\overline{R}X = \{x \in X \mid [x]_R \cap X \neq \emptyset\}$, called lower approximation of X and upper approximation of X with respect to R .

If $\underline{R}X = \overline{R}X$, we say that X is R -definable, otherwise X is said to be rough with respect to R .

Combining the notions of a fuzzy set and rough sets together, Dubois and Prade [21] have defined the notions of a rough fuzzy set and fuzzy rough set. We present below the notion of a rough fuzzy set.

Definition 3.4. Let (U, R) be an approximation space and $U/R = \{X_1, X_2, \dots, X_n\}$. Then for any $X \in F(U)$, $\underline{R}X$ and $\overline{R}X$, the lower and upper approximations of X with respect to R are fuzzy sets in U/R . That is, $\underline{R}X, \overline{R}X : U/R \rightarrow [0, 1]$ such that for all $j = 1, 2, \dots, n$;

$$\underline{R}X(X_j) = \inf_{y \in X_j} \mu_X(y) \text{ and } \overline{R}X(X_j) = \sup_{y \in X_j} \mu_X(y).$$

If $\overline{R}X \neq \underline{R}X$, then the pair $(\underline{R}X, \overline{R}X)$ is called a rough fuzzy set associated with X with respect to R .

3.1. Conventional edge detection techniques

The problem of false edge detection generates thin or thick lines, filters out useless information, by preserving the significant structural properties in a leukaemia image. We can have four conventional edge detection techniques – Sobel, Canny, Robert Cross and Zero-Cross [29], which are applied for medical images.

3.2. Otsu thresholding, statistical methods and histogram analysis

Thresholding is computationally inexpensive and fast. It is one of the oldest segmentation methods and is still widely used in simple applications. Using range values or threshold values, the pixels are classified and clustered using either of the thresholding techniques, like global and local thresholding.

3.2.1. Statistical methods

Different mathematical methods can be applied on the bit-plane to get the enhanced leukaemia image [30, 31]. In the paper, the root mean square error and peak signal noise ratio have to be computed for better interpretation of a leukaemia image.

3.2.2. RMSE and PSNR values

The Root Mean Square Error (RMSE) between an input image and an output image is a measure of the level of information loss. If the size of the input image $i(x, y)$ and the output image $o(x, y)$ is $M \times N$, then the RMSE can be computed by using the formula

$$\text{RMSE} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (i(x, y) - o(x, y))^2 \right]^{1/2}.$$

On the other hand, the Peak Signal-to-Noise Ratio (PSNR) is closely related to RMSE and can be computed from it. But the lower value of RMSE and higher value of PSNR indicate better results.

3.2.3. Histogram analysis

A histogram is a graphical representation of the **distribution** of data. In an image processing context, the histogram of an image normally refers to a histogram of the **pixel intensity values**. This histogram is a graph, showing the number of **pixels** in an image for each different intensity value found in the image. An image histogram is a type of a **histogram** that acts as a **graphical representation** of the **tonal distribution** in a **digital image**. It plots the number of **pixels** for each tonal value. By looking at the histogram for a specific image, a viewer will be able to judge the entire tonal distribution at a glance. A histogram can be applied to review graphically the images resulting in terms of allocation and deviation by representing the gray levels of the image to be enhanced.

3.3. DB index and D index

The Davis-Bouldin (DB) and Dunn (D) indices are two of the basic performance analysis indices. They help in evaluating the efficiency of clustering. The results are dependent on the number of clusters one requires.

3.3.1. Davis-Bouldin (DB) index

The DB index is defined as the ratio of the sum of the within-cluster distance to the between-cluster distance. It is formulated as given in [22] as follows:

$$(1) \quad B = \frac{1}{c} \sum_{i=1}^c \max_{k \neq i} \left\{ \frac{s(v_i) + s(v_k)}{d(v_i, v_k)} \right\} \quad \text{for } 1 < i, k < c.$$

The within-cluster distances, $s(v_i)$ have been formulated independent of any clustering approach. Herein, $d(v_i, v_k)$ is the distance between the centres v_i and v_k of the i -th and the k -th clusters. The aim of this index is to minimize the within cluster distance and maximize the between cluster separation. Therefore, a good clustering procedure must give the value of DB index as low as possible [26], respectively.

3.3.2. Dunn (D) index

Similarly to the DB index, the D index is used for identification of clusters that are compact and separated. It is computed by using

$$(2) \quad D = \min_i \left\{ \min_{k \neq i} \left\{ \frac{d(v_i, v_k)}{\max_l s(v_l)} \right\} \right\} \quad \text{for } 1 < k, i, l < c.$$

It aims at maximizing the between-cluster distance and minimizing the within-cluster distance. Hence, a greater value for D index proves to be more efficient.

4. Methodology

In this section we describe the methodology used in this paper. First we introduce the refined bit plane algorithm in the next subsection.

4.1. Refined bit plane algorithm

START

Step 1. Input image: Read the Leukaemia image in a jpeg format of size (512×512)

Step 2. Convert the RGB image to a gray scale image

Step 3. Apply contrast enhancement using the histogram equalization for better quality of the image

Step 3.1. Perform contrast enhancement by checking the validity of an array

Step 3.2. Histogram normalization using
 $hnorm = hnorm * (\text{variable}(z) / \text{sum}(hnorm));$

Step 4. Bit planes extraction on histogram equalization

Step 4.1. Use a loop to iterate 1:8 planes

Step 4.1.1. Assign bitplanes of the image to $x\{i\}$

Step 5. If (bit plane reconstruction after extraction of the 8 bit planes)

Step 5.1. Assign the best bit plane based on probability as

$B = \text{zeros}(\text{size}(im)) \dots$

Output Image:

Step 5.2. Display the refined best bit plane (enhanced image)

Step 6. Output image: A refined bit plane leukaemia image

STOP

The bit plane algorithm is used to partition the image into 0-7 slices. For example, the sliced value “10011011” in terms of bits is equivalent to 155 in decimal. Pre-processing of leukaemia images is an important step in the process of detection and extraction of the significant features in them [30, 31]. If a bit on the n -th bit plane on an m -bit dataset is set to 1, it contributes a value of $2^{(m-n)}$, otherwise it contributes nothing. In this algorithm contrast enhancement is applied using histogram equalization to improve the quality of an image before slicing it into 0-7 planes.

4.2. Existing algorithms

In this section we present the five uncertainty based clustering techniques to be used in this paper.

4.2.1. Fuzzy C-Means

FCM is an algorithm proposed by Bezdek [18]. In fuzzy clustering (also referred to as soft clustering), the data elements can belong to more than one cluster, a set of membership levels is associated with each element. They indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

Step 1. Assign initial cluster centers or means for c clusters.

Step 2. Calculate the distance d_{ik} between the data objects x_k and centroids v_i using the Euclidean distance formula

$$(3) \quad d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}.$$

Step 3. Generate the fuzzy partition matrix or membership matrix U :

$$(4) \quad \mu_{ik} = \begin{cases} \left\{ \sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)} \right\}^{-1}, & \text{if } d_{ij} > 0; \\ 1, & \text{otherwise.} \end{cases}$$

Step 4. The cluster centroids v_i are calculated using the formula

$$(5) \quad v_i = \frac{\sum_{j=1}^N (\mu_{ij})^m x_j}{\sum_{j=1}^N (\mu_{ij})^m}$$

Step 5. Calculate the new partition matrix by using Steps 2 and 3.

Step 6. If $\|U^{(r)} - U^{(r+1)}\| < \varepsilon$ then stop, else repeat from Step 4.

4.2.2. Rough C-Means

RCM algorithm was introduced by Lingras and West [19] to describe a cluster by its centroid and its lower and upper approximation. The elements in the lower approximation are certain elements and are assigned a higher weight w_{low} than the elements in the upper approximations, which is denoted by w_{up} . An element can belong to two upper approximations. The algorithm is in five steps.

Step 1. Assign initial means v_i for c clusters.

Step 2. Compute μ_{ik} using (4.2.1.2)

Step 3. Let μ_{ik} and μ_{jk} be the maximum and next to the maximum membership values of the object x_k to cluster centroids v_i and v_j .

If $\mu_{ik} - \mu_{jk} < \varepsilon$ then

$x_k \in \overline{BU}_i$ and $x_k \in \overline{BU}_j$, and x_k cannot be a member of any lower approximation.

Else $x_k \in \underline{BU}_i$.

Step 4. Calculate the new cluster means by using

$$v_i = \begin{cases} w_{\text{low}} \frac{\sum_{x_k \in \underline{BU}_i} x_k}{|\underline{BU}_i|} + w_{\text{up}} \frac{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} x_k}{|\overline{BU}_i - \underline{BU}_i|} & \text{if } \underline{BU}_i \neq \phi \text{ and } (\overline{BU}_i - \underline{BU}_i) \neq \phi; \\ \frac{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} x_k}{|\overline{BU}_i - \underline{BU}_i|} & \text{if } \underline{BU}_i = \phi \text{ and } (\overline{BU}_i - \underline{BU}_i) \neq \phi; \\ \frac{\sum_{x_k \in \underline{BU}_i} x_k}{|\underline{BU}_i|} & \text{otherwise.} \end{cases}$$

Step 5. Repeat from Step 2 until a termination condition is met or until there are no more assignments of objects.

4.2.3. Intuitionistic fuzzy C-Means

The IFCM proposed by Ch a i r a and S n e h [25] brings into account the hesitation function that helps in increasing the accuracy of clustering.

Step 1. Assign the initial cluster centers or means for c clusters.

Step 2. Calculate the distance d_{ik} between the data objects x_k and centroids v_i , using formulae (3) and (5) respectively.

Step 3. Generate the fuzzy partition matrix or membership matrix U :

Compute μ_{ik} using the formula (4)

Step 4. Compute the hesitation matrix π using

$$(6) \quad \pi_A(x) = 1 - \mu_A(x) - \frac{1 - \mu_A(x)}{1 + \lambda \mu_A(x)} \quad | \quad x \in X.$$

Step 5. Compute the modified membership matrix U' using

$$\mu'_{ik} = \mu_{ik} + \pi_{ik}$$

Step 6. The cluster centroids are calculated using formula (4).

Step 7. Calculate a new partition matrix by using Steps 2 up to 5.

Step 8. If $\|U^{(r)} - U^{(r+1)}\| < \varepsilon$ then stop, otherwise repeat from Step 4.

4.2.4. Rough fuzzy C-Means

RFCM is an algorithm proposed by M i t r a [32] and M a j i et al. [22]; it combines the concepts of the rough set theory and fuzzy set theory. The concepts of lower and upper approximations in a rough set deals with uncertainty, vagueness and incompleteness, whereas the concept of a membership function in a fuzzy set helps in enhancing and evaluating of overlapping clusters. Algorithm is in five steps.

Step 1. Assign the initial means v_i for c clusters.

Step 2. Compute μ_{ik} using (4).

Step 3. Let μ_{ik} and μ_{jk} be the maximum and next to maximum membership values of the object x_k to cluster centroids v_i and v_j .

If $\mu_{ik} - \mu_{jk} < \varepsilon$ then

$x_k \in \overline{BU}_i$ and $x_k \in \overline{BU}_j$, and x_k cannot be a member of any lower approximation.

Else $x_k \in \underline{BU}_i$

Step 4. Calculate the new cluster means by using

$$v_i = \begin{cases} w_{\text{low}} \frac{\sum_{x_k \in \underline{BU}_i} x_k}{|\underline{BU}_i|} + w_{\text{up}} \frac{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m} & \text{if } \underline{BU}_i \neq \phi \text{ and } (\overline{BU}_i - \underline{BU}_i) \neq \phi; \\ \frac{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m} & \text{if } \underline{BU}_i = \phi \text{ and } (\overline{BU}_i - \underline{BU}_i) \neq \phi; \\ \frac{\sum_{x_k \in \underline{BU}_i} x_k}{|\underline{BU}_i|} & \text{otherwise.} \end{cases}$$

Step 5. Repeat from Step 2 until a termination condition is met or until there are no more assignments of objects.

4.2.5. Rough intuitionistic fuzzy C-Means

The RIFCM, Tripathy et al. [26] uses the concept of rough sets and intuitionistic fuzzy sets.

The steps that are to be followed in this algorithm are eight.

Step 1. Assign the initial means v_i for c clusters by choosing any random c objects as a cluster.

Step 2. Calculate d_{ik} using the Euclidean distance formula (3).

Step 3. Compute U matrix by finding the values of μ_{ik} using (4).

Step 4. Compute π_{ik} using (6).

Step 5. Compute μ'_{ik} using $\mu'_{ik} = \mu_{ik} + \pi_{ik}$ and normalize.

Step 6. Let μ'_{ik} and μ'_{jk} be the maximum and next to maximum membership values of the object x_k to the cluster centroids v_i and v_j :

If $\mu'_{ik} - \mu'_{jk} < \varepsilon$ then

$x_k \in \overline{BU}_i$ and $x_k \in \overline{BU}_j$ and x_k cannot be a member of any lower approximation;

else $x_k \in \underline{BU}_i$.

Step 7. Calculate the new cluster means by using the formula below:

$$(7) \quad v_i = \begin{cases} w_{\text{low}} \frac{\sum_{x_k \in \underline{BU}_i} x_k}{|\underline{BU}_i|} + w_{\text{up}} \frac{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m} & \text{if } |\underline{BU}_i| \neq \phi \text{ and } |\overline{BU}_i - \underline{BU}_i| \neq \phi; \\ \frac{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{BU}_i - \underline{BU}_i} \mu_{ik}^m} & \text{if } |\underline{BU}_i| \neq \phi \text{ and } |\overline{BU}_i - \underline{BU}_i| \neq \phi; \\ \frac{\sum_{x_k \in \underline{BU}_i} x_k}{|\underline{BU}_i|} & \text{else.} \end{cases}$$

Step 8. Repeat from Step 2 until a termination condition is met or until there are no more assignments of objects.

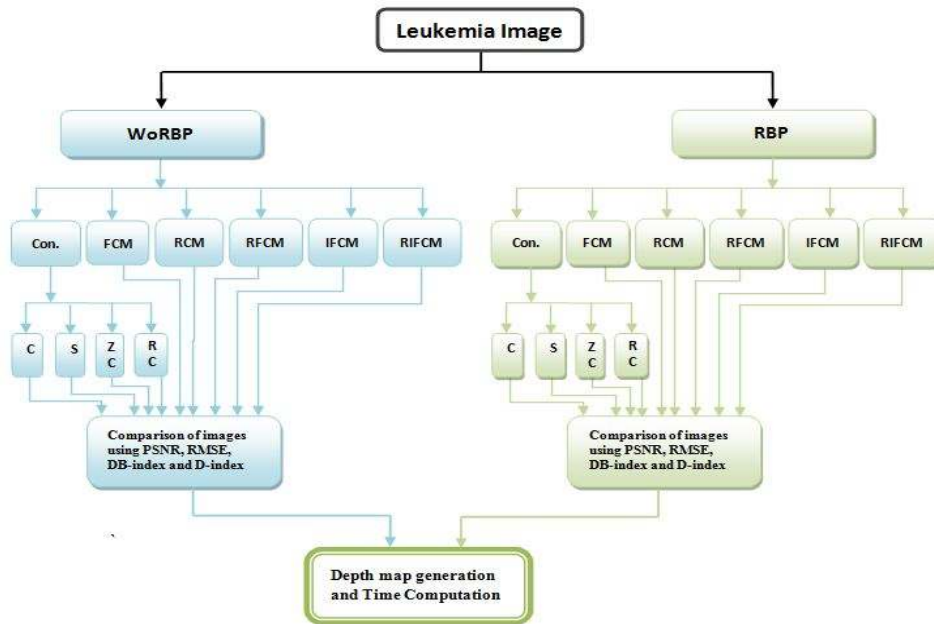


Fig. 1. Block diagram explaining the different steps of the computational procedure

Fig. 1 has four primary steps; the input of leukaemia images, either using the RBP to filter out the noises in them or passing the original image as it is, applying different clustering techniques (both conventional and C-Means), computing different measures of efficiency and the computation of depth of the clustered images, while measuring the time taken for this process. The purpose of using RBP (introduced by us in an earlier paper) on leukaemia images is to reaffirm the necessity of this filtering technique in getting better images while processing the images further. The use of nine clustering techniques is to show the efficiency of C-means clustering techniques over the conventional techniques used for this purpose on one hand, and on the other hand to confirm the superiority of the RIFCM (developed by us in a paper published in 2013) over the other C-Means algorithms in the process, for medical images like those of leukaemia. Finally, the depth computation provides better visibility to assess the quality of the two dimensional images. While generating the depths for the clustered images, we also keep a track

on the time taken. This is another measure to establish the superiority of the RIFCM algorithm, which we observe to be providing the best images and also takes minimum time for the generation of the depth map.

In Table 1 we present the epsilon values when the four conventional clustering techniques are used for segmenting the leukaemia images and the corresponding diagram representation using a bar diagram provided in Fig. 2 to make the comparison easier. The abbreviations WORBP and WRBP are used to represent “Without a Refined Bit Plane” and “With a Refined Bit Plane” respectively. By WORBP we mean that the original image is used without filtering. By WRBP we mean that the image is filtered by using the RBP algorithm before giving it as an input.

The epsilon, which forms the upper bound of the relative error, is extremely useful in determining the convergence tolerances of many iterative numerical algorithms. Image thresholding performs a central role in image segmentation. In Otsu Threshold/Epsilon, we exhaustively search for the epsilon defined as a weighted sum of variances of the two classes: this minimizes the intra-class variance (the variance within the class). Otsu shows that minimizing the intra-class variance is the same as maximizing the inter-class variance. Otsu thresholding/Epsilon is used to make each cluster as tight as possible. Thus it minimizes their overlap. Below we compute the epsilon values for different clustering methods. The lower the epsilon value, the better the clustering method is.

Table 1. Epsilon values for an original leukaemia image and the refined one after applying conventional clustering techniques

Conventional Clustering methods	WORBP	WRBP
Input image	96	80
Canny	79	57
Robert	18	130
Sobel	69	96
Zero-cross	97	97

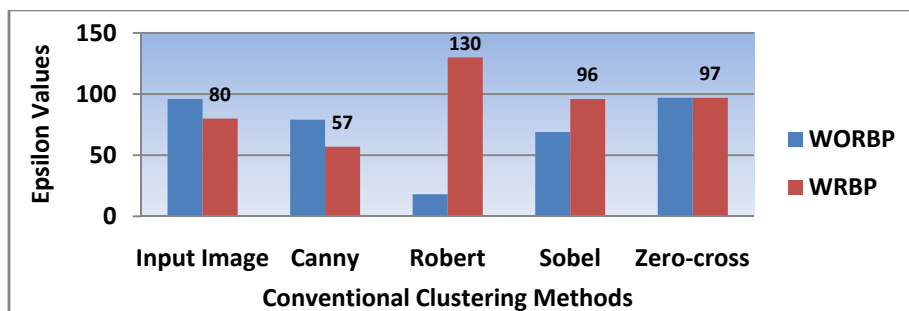


Fig. 2. Diagram representation of the results in Table 2 using a bar diagram

The results of a similar kind of analysis for the five different C-Means clustering algorithms are presented in Table 2 and the corresponding bar diagram is presented in Fig. 3. It can be observed that for all clustering algorithms the epsilon values are lower when RBP is used than when RBP is not used. This is true even in the case of an original image.

Table 2. Epsilon values for the images using different C-Means clustering methods

C-Means Algorithms	WORBP	WRBP
Input Image	96	80
FCM	119	102
RCM	123	102
RFCM	126	103
IFCM	130	110
RIFCM	135	117

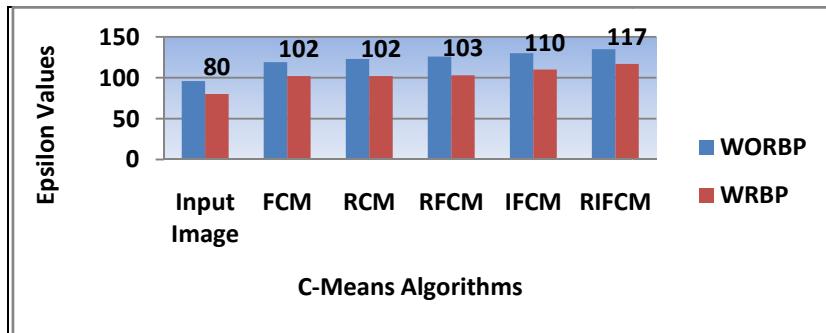


Fig. 3. Diagram representation of the epsilon values obtained in Table 2

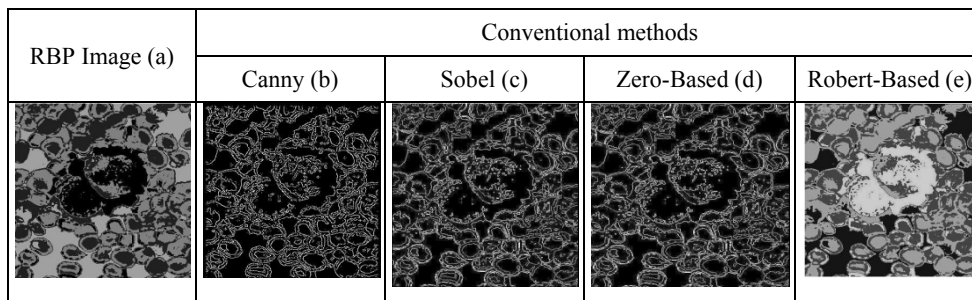


Fig. 4. Original leukaemia image and images after applying conventional methods without RBP

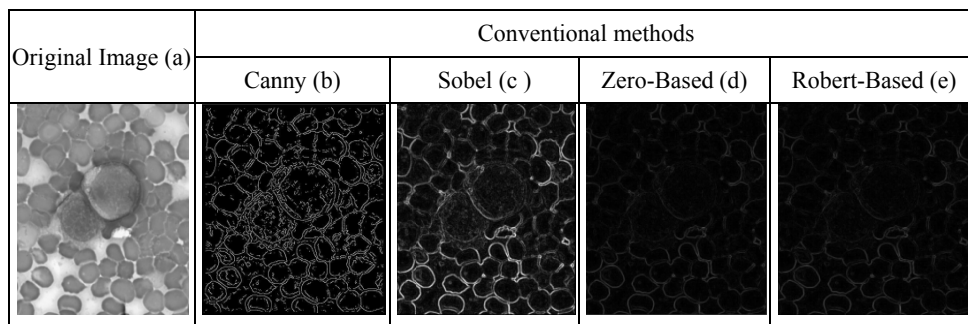


Fig. 5. Leukaemia image after using RBP and the images after applying conventional methods to it

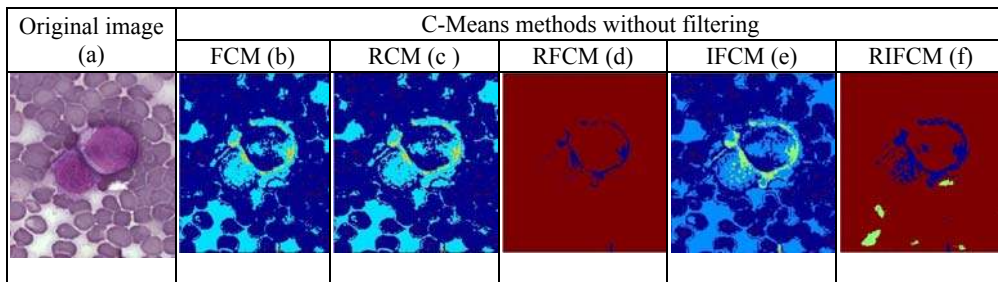


Fig. 6. Original leukaemia image and images after applying C-Means methods

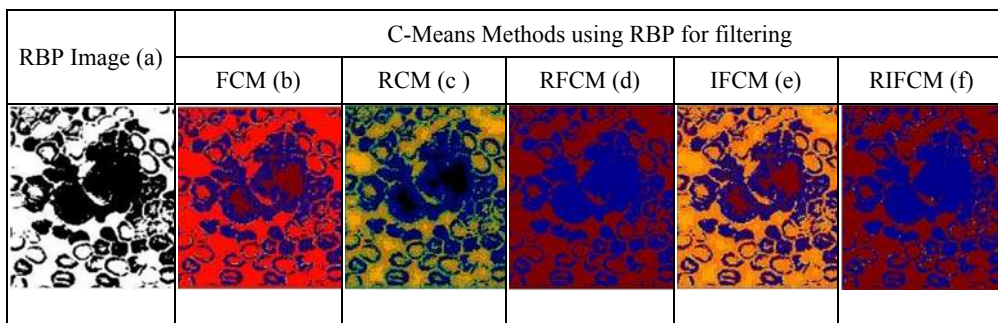


Fig. 7. Leukaemia image after using RBP and the images after applying C-Means methods to it

In the following representation in the form of Table 3, we present the leukaemia images for all the nine methods (four conventional methods and five C-Means methods). In the first column the images were not filtered, while in the second column RBP filtering was applied before clustering.

Table 3 also presents the histogram associated with each of the clustered image. Analysis of these histograms provides the tonal distribution in a leukaemia image. As the histogram plots the number of pixels for each tonal value, by looking at it a doctor can judge the entire tonal distribution at a glance, which determines whether a patient is suffering from cancer or not. It is observed that in this case the RIFCM algorithm also provides a better histogram in the case when an image, filtered by using RBP, is given as an input.

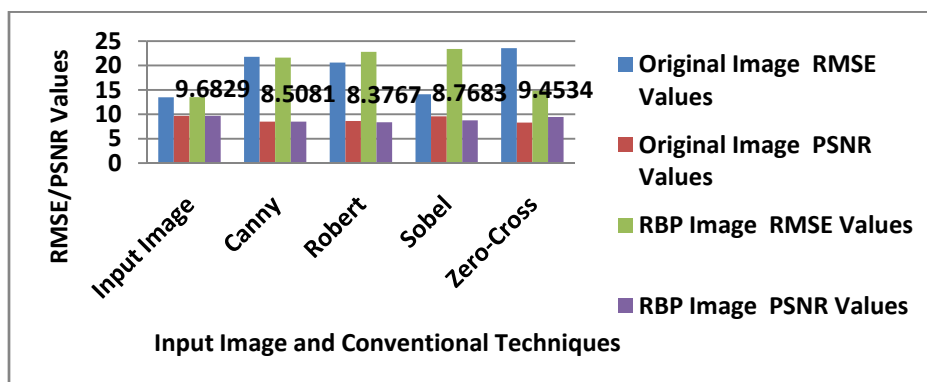


Fig. 8. Graphical representation of the values obtained in Table 4 using a bar diagram

Table 3. Histogram analysis of a leukemia image and images obtained after applying the conventional and the C-Means Clustering methods: without applying RBP, (a); after applying RBP, (b)

Original Image and images after applying different clustering techniques		Histogram Analysis	Original Image applying RBP and images after applying different clustering techniques to it		Histogram Analysis
Original Image			RBP Image		
Canny			Canny		
Sobel			Sobel		
Zero-Cross			Zero-Cross		
Robert-Cross			Robert-Cross		
FCM			FCM		
RCM			RCM		
RFCM			RFCM		
IFCM			IFCM		
RIFCM			RIFCM		

Table 4. RMSE and PSNR values of a leukaemia image when conventional clustering techniques are used: i) the original image ii) the image after applying RBP filtering

Conventional methods	Original image		RBP image	
	RMSE values	PSNR values	RMSE values	PSNR values
Input image	13.47	9.6922	13.52	9.6829
Canny	21.80	8.4886	21.63	8.5081
Robert	20.60	8.6301	22.80	8.3767
Sobel	14.11	9.5761	23.40	8.7683
Zero-Cross	23.57	8.2934	14.82	9.4534

Table 5. RMSE and PSNR values of i) original leukaemia images ii) the image after applying RBP obtained after segmentation using different C-Means algorithms

C-Means & the proposed method	WORBP		WRBP	
	DB-Index	Dunn Index	DB-Index	Dunn Index
FCM	0.0189	18.9015	0.02764	15.7865
RCM	0.0168	45.4763	0.02354	18.9954
RFCM	0.0375	19.6399	0.0157	7.1527
IFCM	0.0179	38.7609	0.0218	17.3894
RIFCM	0.0194	68.9092	0.0159	88.9092

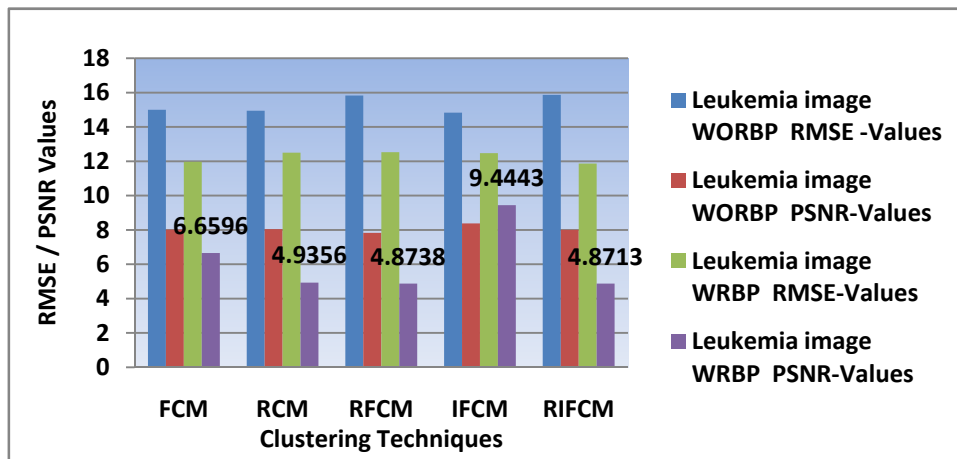


Fig. 9. Graphical representation of the values obtained in Table 5 using a bar diagram

It could be observed from Table 5 and Fig. 9 that there is no uniformity in segmenting the original leukaemia image whereas for the filtered image, using RBP, RIFCM provides the best results in the form of lower values for both the RMSE and PSNR.

Table 6. DB and Dunn Index values for the leukaemia image when clustered using the five C-Means algorithms: i) the original image ii) the image filtered using RBP

Clustering techniques used	Leukaemia image WORBP		Leukaemia image WRBP	
	RMSE -Values	PSNR-Values	RMSE-Values	PSNR-Values
FCM	15.00	8.0139	11.9649	6.6596
RCM	14.9392	8.0491	12.5012	4.9356
RFCM	15.831	7.8334	12.5237	4.8738
IFCM	14.826	8.38	12.4722	9.4443
RIFCM	15.8698	8.0117	11.8652	4.8713

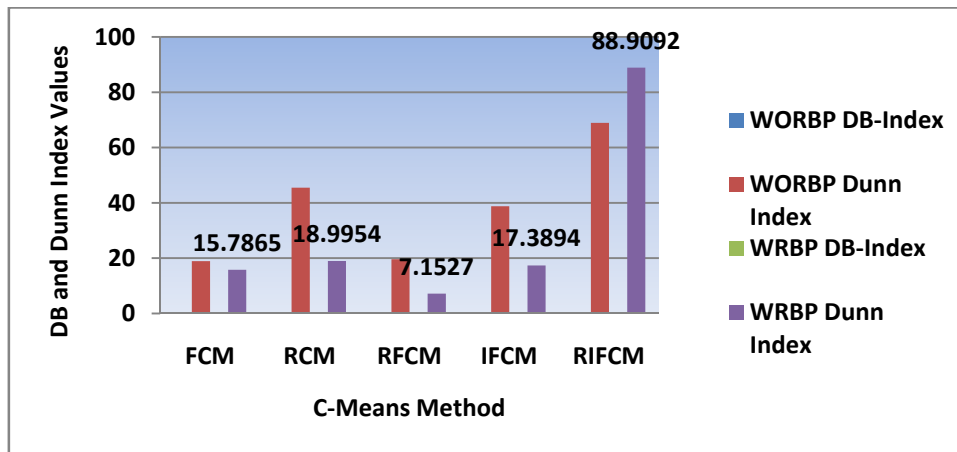


Fig. 10. Graphical representation of the values obtained in Table 6 using a bar diagram

It is clear from Table 6 and Fig. 10 that for RIFCM method both the DB value and D value are the best (minimum for the DB case and maximum for the D case). This observation is true for both cases; i.e., when the original image is taken without applying any filtering technique and when the images are first filtered by using the RBP technique.

Above we have analyzed the efficiencies of the conventional clustering techniques and the C-Means algorithms separately when applied to leukaemia images. Besides, we have given the original image, as well as the image filtered by using RBP in order to establish that the filtered images give better results in all the cases taken separately. The measures were the RMSE and PSNR values and the DB and D indices.

However, when the two Tables 4 and 5 are combined, the results confirm the superiority of WRBPRIFCM for clustering of the images with respect to all other methods considered in this paper, which is visually supported by combining Figs 8 and 9.

5. Depth computation

In this section, we deal with another measure to show the superiority of the RIFCM method when applied to filtered leukaemia images. This is the depth computation of the images, which provides the time required to reconstruct the images. The images obtained after the depth map generation provide a better view and hence assists the doctors in the diagnosis of the disease in a better way. That is why, in this section we deal with this aspect for all the nine clustering algorithms dealt so far. The depth computation of various clustering techniques taking the original image as an input and the images obtained after filtering by using the RBP technique is made. The results are shown in Table 7 and Fig. 11. Here also the output of the RIFCM method yields improved clustering of a leukaemia image and the time of computation time in seconds for this method is the lowest one.

Table 7. Time complexity of Depth Computation of the leukaemia image after applying the nine methods (using RBP or not using RBP)

Clustering approaches used	Original image (time in seconds)	RBP filtered image (time in seconds)
Input image	35.84	28.49
Canny	27.64	26.58
Robert	29.00	22.52
Sobel	30.56	27.54
Zero-cross	29.54	26.14
FCM	24.54	21.19
RCM	26.75	24.47
RFCM	23.13	22.58
IFCM	25.33	22.78
RIFCM	20.11	18.14

In Fig. 11 the first group of bars corresponds to the case when the original images and the second group correspond to the case when the filtered images using the RBP technique are given as an input.

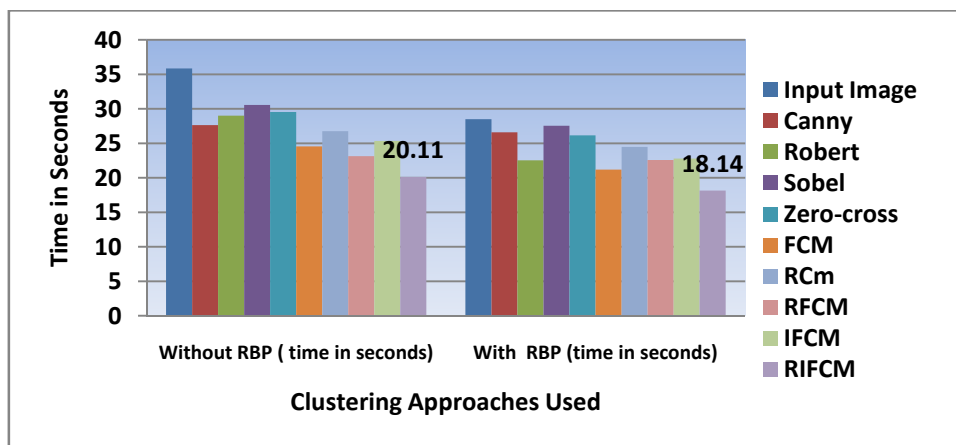


Fig. 11. A graph representing the time complexity in seconds of Depth Computation for the leukaemia image after applying the nine methods (using RBP or not using RBP)

6. Conclusion and future work

The primary objective of the work in this paper was to provide a method which is likely to assist the doctors by processing the leukaemia images to get a better idea in determining whether a patient is suffering from cancer or not. We succeeded in it by considering nine clustering techniques (both conventional and C-Means), such that the outputs are better images. Our approach also reaffirms the utility of RBP in filtering the images leading to better analysis and it also establishes that the RIFCM hybrid clustering algorithm when applied to filtered leukaemia images provides the best results among them. Our approach will assist doctors to analyze the stage of the cancer and proceed in a further treatment. In order to measure the efficiency of the methods included in our study, we used four measures in the form of RMSE, PSNR, epsilon computation, DB and D indices and the time required in the depth map generation. Besides, we have applied histogram analysis to provide easy interpretation of leukaemia images.

References

1. Mohapatra, S., S. S. Samanta, D. Patra, S. Satpathi. Fuzzy Based Blood Image Segmentation for Automated Leukemia Detection. – In: Proc. of International Conference on Devices and Communications, 2011, pp. 1-5.
2. Swarnalatha, P., B. K. Tripathy. Depth Computation Using Bit Plane with Clustering Techniques for Satellite Images. – International Journal of Earth Sciences and Engineering, Vol. 6, No 6(01), 2013, pp. 1541-1553.
3. Swarnalatha, P., B. K. Tripathy, S. Prabu, R. Ramakrishnan, M. S. Moorthi. Depth Reconstruction Using Geometric Correction with Anaglyph Approach for Satellite Imagery. – In: Proc. of International Conference on Advances in Communication, Network, and Computing, CNC, Elsevier, 2014, pp. 506-515.
4. Swarnalatha, P., M. Kota, N. R. Resu, G. Srivasanth. Automated Assessment Tool for the Depth of Pipe Deterioration. – In: Proc. of IEEE Advance Computing Conference, 2009, , pp. 721-724. ISBN 978-1-4244-2928-8.
5. Swarnalatha, P., B. K. Tripathy. A Centroid Model for the Depth Assessment of Images Using Rough Fuzzy Set Techniques. – International Journal of Intelligent Systems and Applications, Vol. 1, 2012, pp. 20-26.
6. Chan, Y. K., M. H. Tsai, D. C. Huang, Z. H. Zheng, K. D. Hung. Leukocyte Nucleus Segmentation and Nucleus Lobe Counting. Bio Med Central, Ltd., 2010.
7. Mecca, G., S. Raunich, A. Rappalardo. A New Algorithm for Clustering Search Results. – Data and Knowledge Engineering, Vol. 62, 2007, pp. 504-522.
8. Xu, R., D. Wunsch. Clustering Algorithm in Biomedical Research: A Review. – IEEE Reviews in Biomedical Engineering, Vol. 3, 2010, pp. 120-154.
9. Valafar, F. Pattern Recognition Techniques in Microarray Data Analysis: A Survey. – Annals of New York Academy of Sciences, Vol. 980, 2002, pp. 41-64.
10. Jain, A. K., R. C. Dubes. Algorithms for Clustering Data. Upper Saddle River, NJ, USA, Prentice Hall, 1988.
11. Bigus, J. P. Data Mining with Neural Networks. McGraw-Hill, 1996.
12. Hazlyna, H. N., M. Y. Mashor, N. R. Mokhtar, S. A. N. Aimi, H. Rosline, R. A. A. Raof, M. K. Osman. Comparison of Acute Leukemia Image Segmentation Using HSI and RGB Color Space. – In: Proc. of 10th International Conference on Information Science, Signal Processing and Their Applications ISSPA-2010, pp. 749-752.

13. Mohapatra, S., D. Patra, K. Kumar. Fast Leukocyte Image Segmentation Using Shadowed Sets. – *International Journal of Computational Biology and Drug Design*, Vol. **5**, 2012, No 1, pp. 49-65.
14. Fatichah, C., M. L. Tangel, M. R Widyanto, F. Dong, K. Hirota. Interest-Based Ordering for Fuzzy Morphology on White Blood Cell Image Segmentation, 2012.
15. Pandian, A. P., S. N. Sivanandam. Hybrid Algorithm for Lossless Image Compression Using Simple Selective Scan Order with Refined Bit Plane Slicing – *Journal of Computer Science*, Vol. **8**, 2012, Issue 8, pp. 1338-1345.
16. Tou, T., R. C. Gonzalez. *Pattern Recognition Principles*. London, Addison-Wesley, 1974.
17. Xiong, H., J. J. Wu, J. Chen. K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective. – *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, Vol. **39**, 2009, pp. 318-331.
18. Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA, Kluwer Academic, 1981.
19. Lingras, P., C. West. Interval Set Clustering of Web Users with Rough K-Means. – *Journal of Intelligent Information Systems*, Vol. **23**, 2004, pp. 5-16.
20. Pawlak, Z. Rough Sets. – *International Journal of Computer and Information Sciences*, Vol. **11**, 1982, pp. 341-356.
21. Dubois, D., H. Prade. Rough Fuzzy Sets Model. – *International Journal of General Systems*, Vol. **46**, 1990, No 1, pp. 191-208.
22. Maji, P., S. K. Pal. RFCM: A Hybrid Clustering Algorithm Using Rough and Fuzzy Sets. – *Fundamenta Informaticae*, Vol. **80**, 2007, No 4, pp. 475-496.
23. Swarnalatha, P., B. K. Tripathy, N. P. Ladda, D. Ghosh. Cluster Analysis Using Hybrid Soft Computing Techniques. – In: *Proc. of International Conference on Advances in Communication, Network, and Computing, CNC*, Elsevier, 2014, pp. 516-524.
24. Atanassov, K. T. Intuitionistic Fuzzy Sets. – *Fuzzy Sets and Systems*, Vol. **20**, 1986, No 1, pp. 87-96.
25. Chaira, T., A. Sneh. A Novel Intuitionistic Fuzzy Approach for Tumor/Hemorrhage Detection in Medical Images. – *Journal of Scientific & Industrial Research*, Vol. **70**, 2011, pp. 427-434.
26. Tripathy, B. K., Rohan Bhargava, Anurag Tripathy, Rajkamal Dhull, Ekta Verma, P. Swarnalatha. Rough Intuitionistic Fuzzy C-Means Algorithm and a Comparative Analysis. – In: *Proceedings of ACM Compute-2013*, VIT University, 21-22 August 2013.
27. Swarnalatha, P., B. K. Tripathy. A Comparative Study of RIFCM with Other Related Algorithms from Their Suitability in Analysis of Satellite Images Using Other Supporting Techniques. – *Kybernetes*, Emerald Publications, Vol. **43**, 2014, No 1, pp. 53-81.
28. Zadeh, L. A. Fuzzy Sets. – *Inform. Control*, Vol. **8**, 1965, pp. 338-353.
29. Raman, M., H. Aggarwal. Study and Comparison of Various Image Edge Detection Techniques. – *International Journal of Image Processing*, Vol. **3**, 2002, Issue 1.
30. Ersin, G. Adaptive Wiener-Turbo Systems with JPEG & Refined Bit Plane Compressions in Image Transmission. – *Turk Journal Electrical Engineering & Comp. Science*, Vol. **19**, 2011, pp. 141-155.
31. Ashok, Kumar T., S. Priya, Varghese Paul. Automatic Feature Detection in Human Retinal Imagery Using Bitplane Slicing and Mathematical Morphology. – *European Journal of Scientific Research*, Vol. **80**, 2012, pp. 57-67.
32. Mitra, S. An Evolutionary Rough Partitive Clustering. – *Pattern Recognition Letters*, Vol. **25**, 2004, pp. 1439-1449.