

Enhanced Support Vector Machine Based Leukaemia Cancer Classification



T. Preethi, D. Maheswari

Abstract: Leukaemia is a blood cancer that is characterized by the bone marrow's unregulated and irregular generation of white blood cells (leukocytes) in the blood. By testing the microscopic blood cell images, diseases can be detected and its early diagnosis can be done. In the current work, at smart decision support system utilizing microscopic images is proposed to make a diagnosis of acute lymphoblastic leukaemia. Identification through images is a quick and inexpensive technique since there is no equipment is specially needed for lab testing. The improved Bare-Bones Particle Swarm Optimization algorithm is used to identify the important distinguishing features of blast cells and the healthy cells to help in carry out the acute lymphoblastic leukaemia classification with efficiency. To avoid the poor accuracy, the modified median filtering has been introduced in this novel, which helps in removing the noise from the images and maintains the edge which is used to improve the identification of leukocytes and lymphocytes during the segmentation process. Enhanced linear contrast stretching is introduced in image enhancement for enriching the image. In the next step, feature extraction is carried out through the second order statistical features. Bare-bones with Adaptive Bat Optimization (BBABO) is used for feature selection. At last, the classification is carried out by using the combined Fuzzy Neural Network (FNN) and here the Enhanced Support Vector Machine (ESVM) is a classifier called as FNESVM.

KEYWORDS: Leukaemia, white blood cells, lymphoblastic leukaemia, noise removal, Improved median filtering, Adaptive Bat Optimization, Fuzzy Neural Network, Enhanced Support Vector Machine.

I. INTRODUCTION

Leukaemia is a type of cancer belonging to the bone-marrow or blood characterized by an uneven growth of undeveloped white blood corpuscles called as "blasts"[1]. Detection of leukaemia depends, the count of white blood cell which is higher compared to earlier state in addition to undeveloped blast (i.e. lymphoid or myeloid) cells accompanied with reduced number of platelets and neutrophil. The presence of excessive amount of blast-cells in marginal blood proves to be a significant symptom of leukaemia cancer [2-3]. Therefore, haematologists constantly check the blood tissues under optical microscope for performing the classification and also the detection of blast-cells with accuracy.

The diagnosis of the cancer in the early stages can be of great help in reducing the death rate in human beings. Manual morphological analysis may be affected by various probable drawbacks (e.g. unstandardized precision and heavy dependence on medical experts' knowledge and capability) [4, 5, 6].

To overcome these difficulties, several automated diagnosis techniques have been introduced recently to identify the features of healthy and blast cells, which is a critical factor. Even though several research works exist in theseparation and identification of the cytoplasm and nucleus or the nuclei is engaged with the segmentation approach, very less research have been carried out on choosing the important distinguishing features from the segmented regions to make the next acute lymphoblastic leukaemia (ALL) diagnosis process to be supremely effective[7].

In the already available system, two improved Bare-bones particle swarm optimization (BBPSO) algorithms have been introduced for identifying the most important differentiating features of healthy and blast cells to help in carrying out acute lymphoblastic leukaemia classification efficiently. There are multiple factors that have an adverse effect on the quality of images, and they include false background, salt or pepper noise, and low contrast. This leads to poor accuracy, and susceptibility to errors owing to different human factors [8].

In order to resolve those issues encountered in this research work, improved median filtering has been introduced, which eliminates the noise from the images by preserving the edges, and there by aids in improving the identification of the leukocytes and lymphocytes during the segmentation process [9]. Improved linear contrast stretching is suggested for image enhancement to boost the image. After this, feature extraction is carried out through the second order statistical features. In the proposed research work, rather than making use of the BBPSO algorithm, Bare-bones with Adaptive Bat Optimization (BBABO) is introduced for feature selection process. At last, the classification task is carried out by integrating both the Fuzzy Neural Network (FNN) and Enhanced Support Vector Machine (ESVM) classifiers [10] called as FNESVM.

II. LITERATURE REVIEW

Khobragade et al [10] demonstrated a technique of leukaemia detection in patients from testing the microscopic white blood cell images. The focus is on the variations in the geometry of cells and statistical parameters such as mean and standard deviation which differentiates the white blood cells from other blood components employing processing tools including MATLAB and Lab VIEW. The steps of images processing such as image enhancement, image segmentation and feature extraction are used on microscopic images.

Manuscript published on November 30, 2019.

* Correspondence Author

T. Preethi*, Research Scholar, Rathnavel Subramaniam College of Arts and Science, Coimbatore, Tamil Nadu. Email: preethi.thamilanban@gmail.com

D. Maheswari, Head & Research Coordinator, School of Computer Studies- PG, Rathnavel Subramaniam College of Arts and Science, Coimbatore, Tamil Nadu. Email: maheswari@rvsgroup.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Kumar et al [11] to detect the acute leukaemia, an algorithm was introduced for automated image based detection system. The technique devised makes use of the fundamental enhancement, morphology, filtering and segmenting approach for the extraction of the area of interest is used by the algorithm termed as k-means clustering. It is with the exactness of 92.8% and then it is hardened with Nearest Neighbour (kNN) and Naïve Bayes Classifier using 60 samples of data sets.

Sajjad et al [12] studied about an effective framework for localizing the WBCs present within microscopic blood smear images employing a multi-class ensemble classification scheme. In the novel framework, the nuclei are initially segmented, which is then followed by extracting the features like texture, statistical, and wavelet features. At last, the identified WBCs are categorized into five groups, which include basophil, eosinophil, neutrophil, lymphocyte, and monocyte. The results of experiments carried out on a general (non-synthetic) standard database corroborate the efficiency and resourcefulness of the proposed system contrary to state-of-the-art approaches.

Hiremath et al [13] every segment is helpful in identifying and classifying the white blood cells by using the techniques like color based segmentation and the geometric features. The results from the experiments and the manual results achieved by the pathologist are compared and the efficiency of the proposed technique is proven.

Choudhary et al [14] presented an image processing approach for leukaemia detection in a human blood sample. The novel work surpasses the issue of k-means clustering and thresholding technique by making use of the image enhancement approaches and few arithmetic operations for the nucleus segmentation from the white blood cells. The process of segmentation based on LAB color space (luminosity, chromaticity layer A and chromaticity layer B) color space will be utilized for removing the white blood cells (WBC) out of the background. The segmented image is utilized for calculating the shape based feature of the nucleus of the WBCs. K-NN classifier has been used for classifying the blast cells from the common lymphocyte cells.

Mohapatra and Patra [15] had said about the features which is extracted based on shapes i.e., hausdorff dimension and contour signature is used to classify the lymphocytic cell nucleus. The classification is performed with the help of Support Vector Machine (SVM). An overall of 108 blood sample images are taken and measured for feature extraction. This result is analysed and compared with the extraction from a haematologist.

Goutam and Sailaja, [16] studied about four important stages, which include the pre-processing, segmentation, feature extraction and classification correspondingly. This system framework comprises of an easy and known approach that includes Local Directional path, K-mean clustering and support vector machine correspondingly. The classifier classifies the state of a patient to be normal or abnormal. The entire system performance is assessed with the help of the parameters defined including sensitivity, specificity, f-measure, and precision that is helpful in computing the accuracy. Nearly 90 microscopic blood

images were tested and the accuracy of 98% is achieved in the proposed work.

Mohapatra, et al [17] suggested a quantitative microscopic technique to differentiate lymphoblast from lymphocytes in microscopic blood image and the samples of bone marrow. This differentiation helps to develop a computer-assisted ALL viewing. By using the stained blood films of microscopic images the automatic identification of lymphoblast is achieved with the support of image segmentation, feature extraction, and classification. Accurate diagnosis of ALL is accomplished by using the improved segmentation technique, predominant features and an ensemble classifier.

III. PROPOSED METHODOLOGY

This section explains about the architecture of the novel Leukemic diagnosis and ALL classification system. It comprises of six components, where the first one involves marker-controlled watershed segmentation based White Blood Cell (WBC) recognition from the blood smear images and second one includes pre-processing the detected WBC based on improved median filtering for noise elimination and Enhanced linear contrast stretching based image enhancement. The third one includes nucleus and cytoplasm isolation employing stimulating discriminate measure (SDM)-based clustering algorithm from the pre-processed WBC. The fourth one involves feature extraction employing second order statistical features, whereas fifth one includes feature selection employing Bare-bones with Adaptive Bat Optimization (BBABO). And finally, the sixth one includes classification of healthy and blasted cells by integrating the both Fuzzy Neural Network (FNN) and Enhanced Support Vector Machine (ESVM) called as FNESVM. The overall architecture of the novel work is illustrated in figure 1.

As per this figure, the process starts with the White blood cell segmentation from the blood smear image employing marker-controlled watershed segmentation algorithm.

A. White blood cell identification using marker controlled watershed segmentation

Isolating the touching objects in an image is one among the more tedious image processing operations. The watershed transform is frequently used for this problem. In this research work, watershed transform is utilized for identifying the White blood cells (WBC) from the blood smear image, and the watershed transform computes the “catchments basins” and “watershed ridge lines” in an image by considering it as a surface where the lightly illuminated pixels are high and darker pixels are low. Over segmentation is one of the most significant disadvantages faced by the watershed transform. The common means of pre-defining the number and approximate position of the regions rendered by the watersheds approach lies in the transformation of the homogeneity of the function to which the algorithm is used. This change is performed through a mathematical morphology operation, geodesic reconstruction, by which the function is transformed such that the minima can be enforced by an external function (which is the marker function).

Every catchment basin with no marking get filled by the morphological restoration and therefore modified into non minima plateaus, which will not generate unique regions during the computation of the final watersheds. Segmentation employing the watershed transforms operates well when you can find, or “mark,” foreground objects and background locations.

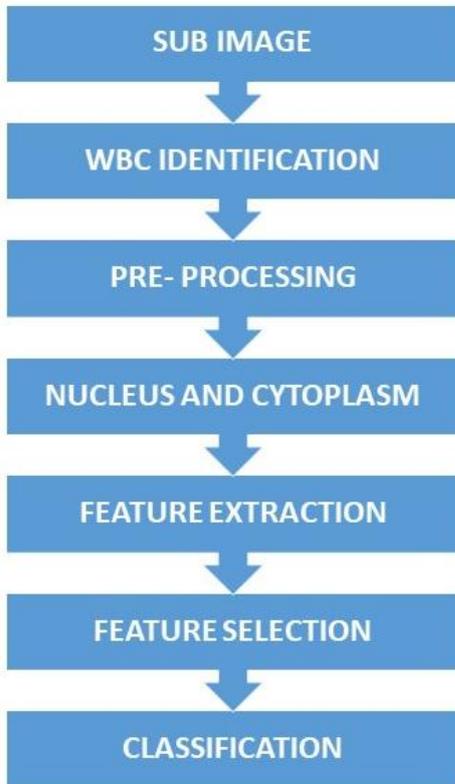


Figure1. Architecture of the proposed work

Marker-controlled watershed Segmentation accepts the fundamental process:

1. The functions of segmentation is computed. This forms an image whose dark areas are the objects that are to be segmented.
2. Foreground markers are computed. These include the linked blobs of pixels present inside each one of the objects.
3. Background markers are computed. These include pixels, which are not a segment of any object.
4. The segmentation function is changed such that it just has a minimum at the foreground and background marker locations.
5. The watershed transform of the modified segmentation function is computed.

Marker controlled watershed segmentation: In this step, the background objects are marked using a marker. Several procedures could be used here to get the foreground markers that need to connect with the pixels of blobs present in the foreground objects. The technique that is utilized involves the morphological schemes known as “opening-by-reconstruction” and “closing by- reconstruction” for “cleaning” the image to produce maximum within every object, were the object can locate employing bioregional max. After this, the background markers are computed. And then the background is computed. In the cleaned-up image, the dark pixels are present in the background, and therefore

a thresholding operation can be used for marking the background objects. After this, the watershed transform of the segmentation function is computed. The function imimposemin can be helpful in modifying an image such that it exhibits regional minima just in few necessary locations. At this stage, imimposemin can be used for modifying the gradient magnitude image.

B. White Improved Filtering For Noise Removal

Once the WBC is identified, noise is eliminated from the segmented WBC for an efficient classification employing Improved Median Filtering.

Median filtering Theory

The median filter is a nonlinear signal processing technology that depends on statistics. The noisy value of the digital image or the sequence is substituted with the median value computed of the neighbourhood (mask). Then the pixels belonging to the mask are sorted according to their grey levels, and the median value of the group is then stored to substitute the noisy value. The median filtering output is $g(x, y) = med\{ f(x - i, y - j), i, j \in W \}$, where $f(x, y)$, $g(x, y)$ refers to the actual image and the output image correspondingly, W indicates the two-dimensional mask: $n \times n$ is the mask size (where n is generally odd) like 3×3 , 5×5 , and etc.; the mask can have linear, square, circular, cross shapes and more.

- *Improved Median Filtering Algorithm (Improvement of the Filtering Mask):* The filtering mask is generally of $n \times n$ square or cross. Going by the symmetry of the mask, n is usually odd. The smaller the mask size is, the better would be the image details that would-be reconstructed, and the less efficient would be the performance of the noise elimination; the greater the mask size is, the less would be retention of the image details, and the more efficient would be the noise elimination performance. In order to resolve the problem, the adaptive filtering algorithm is introduced. During the filtering process, the mask can be adaptively resized based on the noise levels of the mask.
- *Improvements of the Median Algorithm:* As the average filter exhibits a better presentation for filter out random noises, the median filter is combined with the average filter to a particular size of the filter mask. The novel technique can eliminate the noises and helps in retaining the image details in a better way.

Improving information:

For the normal image, neighbouring pixels exhibit a much better association. The grey value of every pixel is quite near to neighbouring pixels, and the edge pixels depict the same characteristic also. In case the value of a pixel is higher or lesser compared to the value in the neighbourhood, the pixel is corrupted by the blare; else, the pixel becomes an existing pixel. During the noise elimination procedure, every pixel is checked in sequence, and in case the value of a pixel is higher than the average rate in the mask, then it is shown that the pixel is corrupted by the noise and it is substituted with the median rate of the mask; else, the actual value of the pixel is unmodified.

This technique not just helps in reducing the time taken in computation, but also maintains the details of the image to the maximum extent possible. The actual value of the pixel is substituted with the median value in the mask, and during the next process of computation the average value may reap the best advantage out of the new rate of the pixel.

This generates an iterative process; it not just reduces the time complexity, but also helps in improving the noise-elimination effect commendably [18].

Fast calculation of median value:

The difficulty of the algorithm is primarily determined by the calculation of the median rate on the above mentioned steps. The research work presents the arithmetical histogram to boost the search speed of the median value.

A. Image enhancement using enhanced linear contrast stretching

Enhancement signifies the changes made to an image to modify the effect on the viewer. Usually, enhancement deforms the actual digital values; hence enhancement is not completed till the restoration processes are done with. There is a huge effect of contrast ratio on improving the power and detection capacity of images. Methods used for boosting the image contrast are among the most elaborately employed enhancement processes. For the generation of an image having the optimum contrast ratio, it is vital to make use of the complete range of brightness of the display medium, which is usually film.

- **Linear Contrast Stretch:** The contrast stretching linearly increases the actual digital values of the noise eliminated WBC image into a brand new distribution. Also, linear contrast enhancement also renders small changes within the data to be more apparent. These kinds of enhancements are best used for images having Gaussian or near-Gaussian histograms, implying that, all the brightness values stay within a short range of the histogram and just one mode is shown. There are three techniques of linear contrast enhancement: Contrast stretching is used on the image for stretching the histogram to complete the entire dynamic range of the image [19]. Two well-known kinds of contrast stretching approaches include the basic stretching contrast and end-in-search. Basic stretching contrast operates well on the image where all the pixels are focused in one segment of the histogram, for instance, in the middle. In addition, contrast stretching is helpful in getting over the shortages or excessive light during shooting, increasing the distribution of pixel-grey values, where the images are generally classified into: low contrast, fine contrast or normal contrast, and high contrast. The image having low contrast is defined by mostly its image compositions of brightness or darkness. The histogram exhibits some level of greyness in groups combined. In case the pixel grouping is present on the left, then the image gets to be dark and vice versa. The image having low quality can be improved using the contrast stretching operation. The contrast stretching algorithm is as given:

1. Get the lower bound of pixel grouping through the scanning of the histogram of the least grey scale to the highest grey scale value (0 to 255) to generate the first pixel, which exceeds the predetermined threshold value.

2. Compute the upper border of pixel grouping through the scanning of the histogram of the highest grey scale value to the least value of the predetermined second threshold value.

3. The pixels lesser than the first threshold value are set with a value of 0, whereas the pixels greater than the second threshold value are 255.

4. The pixels between the first threshold value and the second scaled threshold value to meet the entire range of grey scale values (0 to 255) is given by the mathematical expression in (3).

$$s = \frac{r - r_{\min}}{r_{\max} - r_{\min}} \times 255 \quad (3)$$

Where r refers to the grey scale value of the actual image, s indicates the new grey scale value, the lowest grey scale r_{\min} value of the pixel group, the highest grey scale r_{\max} of the pixel.

- **Piecewise Linear Contrast Stretch Based on Unsharp Masking (PLCSUM):** In this article, piecewise Linear Contrast Stretch Based on Unsharp Masking (PLCSUM) technique is introduced for enhancement of the leather image. Unsharp masking (UM) is basically an image manipulation approach. The Unsharp masking approach can of very much use in improving the detailed appearance using minute enhancements of edge contrast of an image. Generally, Unsharp mask is utilized for sharpening an image where this will be useful in affirming the image texture and image details. The traditional unsharp masking algorithm can be expressed as (5).

$$z = n + \gamma(m - n) \quad (6)$$

Where m indicates the input image, n refers to the result of the process employing a linear low-pass filter, whereas \hat{U} refers to the gain with ($\hat{U} > 0$) that is actually the real scaling factor. Signal $d = m - n$ frequently amplifies ($\hat{U} > 1$) to maximize the sharpness. The signals comprises of image details, noise, over-shoot and under-shoots in a region of sharp edge resulting due to smoothening edges.

If enhancement of a noise cannot be performed, then enhancements of over-shoot and under-shoots are achieved with the help of a visually undesirable halo effect. For this, a filter with no sensitivity to noise and with smoother sharp edge is necessary. The earlier research works had made use of various popular filters, comprising of cubic filters and edge preserving filters replacing the linear low-pass filters. But, both approaches experience few limitations. Cubic filter does not exhibit sensitivity to noise, whereas edge preserving filter does not exhibit smoother and sharp edges. Therefore in this research work, adaptive gain control is proposed. Fig. 1 illustrates block diagram of the conventional unsharp masking algorithm, utilized as the base for the unsharp masking algorithm introduced in this research work.

An unsharp masking framework is designed for leather image enhancement. This framework suggests generalizing the unsharp masking algorithm integrated with the functions of contrast stretching techniques. This work will present an unsharp masking framework designed for leather image enhancement.

This framework depends on generalizing the unsharp masking algorithm integrated with the functions of adaptive contrast stretching on halo effect problems resolved employing an edge preserve filter. In this research work, the concept of enhancement and sharpening makes use of another process, which uses adaptive contrast stretching algorithm [17] and the output is known as $w(y)$.

The processing of the image details are done with $g(d) = \hat{U}(d) \odot d$, where d refers to an adaptive gain control and it is a function of the amplitude of the detail signal of d . The final outcome of the algorithm is expressed as (6)

$$u = w(y) + [\gamma(d) + d] \quad (7)$$

After the image is enhanced, it is required to isolate the nucleus-cytoplasm from the enhanced WBC image employing stimulating discriminate measure (SDM)-based clustering algorithm.

C. Nucleus- Cytoplasm Separation Using Stimulating Discriminate Measure (SDM) - Based Clustering, Segmentation of Nucleus and Cytoplasm for Leucocytes

The French- American- British classification systems perform the classification of ALL into three subtypes (L1–L3) in accordance with the morphology observed on the nucleus and cytoplasm. Pastel blue and non-granular cytoplasm with closed and clumped nucleus chromatin is generally found in mature lymphocytes. As for the blasts cells (e.g. different subtypes of ALL), differences in terms of the nucleus to cytoplasm ratio, presence of nucleoli and vacuoles, nucleus and cytoplasm colour, in addition to the chromatin patterns are found. Hence, differentiation of nucleus from cytoplasm and the properties of nucleus and cytoplasm have considerable roles to play in accurately diagnosing the normal and abnormal lymphocytes. Also, Enhancement of nucleus-cytoplasm segmentation is one of the most important challenges, which needs the most laborious research attempts [20].

The novel SDM-based clustering algorithm is used on each one of the detected WBC, with the aim of isolating the nucleus and cytoplasm for disease identification. In this research work, SDM is incorporated into the Genetic Algorithm (GA) to boost the FCM algorithm in isolating nucleus and cytoplasm from the lymphocyte/lymphoblast images acquired from ALL-IDB2. On the whole, the segmentation of 180 images having 60 normal (lymphocyte) and 120 abnormal (lymphoblast) are done in the experiment. The proposed clustering algorithm is carried out on the L^* component of the CIELAB colour space as the L^* component is capable of showing more variations between the nucleus and cytoplasm, in which the nucleus is usually darker because of the presence of chromatin while cytoplasm is considerably brighter. Even though the brightness across images differs, the brightness in a specific image during the process of clustering generates a distinguishing factor between nucleus and cytoplasm.

The novel SDM-based clustering algorithm is aimed at improving the segmentation potential of traditional FCM. The algorithm begins with random initializing the population, P , comprising of chromosomes, S_i , where $i = 1, 2, \dots, k$, which denotes the threshold value of three clusters: nucleus, cytoplasm, and the background. In the initialization

step, one among the chromosomes, S_m , is acquired in the form of a seed from the converged solution of FCM to speed up the optimization process, where $S_m \sim P$. As per the threshold value signified by every chromosome, all the pixels in the actual image are classified into three clusters, i.e., A, B, and C, which denotes the clusters of cytoplasm, nucleus, and the background, correspondingly.

In this scenario, every pixel indicates a data sample present in a cluster, and a pixel can just be a member of one cluster at one point of time. Once the pixels are separated, the next step involves chromosome evaluation, in which the chromosome fitness, $F(S_i)$, is computed on the basis of SBSDM and SWSDM.

The objective to get a smaller SWSDM and bigger SBSDM, which specifies a greater

Degree of similarity for within-cluster evaluation and greater isolation between clusters, correspondingly. As stated earlier, there are scenarios where the pixel intensity of nucleus and cytoplasm are very near to one another; hence indicating a much better degree of difficulty in isolating both the clusters. In this condition, two constraints are utilized for helping in the process of segmentation: (i) the nucleus/cytoplasm area must not be lesser than 10% of the respective cytoplasm/nucleus region; (ii) the background area must not be greater than the region enclosing the entire membrane (nucleus + cytoplasm). In case the constraints are not met, a penalty value, $_$, is used to improve $F(S_i)$.

Once the evaluation of $F(S_i)$ is done, a random universal sampling method is helpful in eliminating the bias while selecting the chromosomes for reproduction purposes. Single-point crossover and mutation are utilized for generating new off springs with the probability rates of 0.7 and 0.3, correspondingly. The newly produced offspring are utilized for dividing the pixels into individual clusters (i.e. nucleus, cytoplasm, and background) and further assessed using the fitness function, $F(S_i)$. Afterwards, having a generation gap of 0.9, of spring and parent solutions are then ranked and combined into the new generation. On the basis of multiple trials, the GA is capable of converging to a good differentiation between nucleus and cytoplasm when the maximum number of generations is fixed to 100. Hence, the processes of evaluation, crossover, and mutation are performed again until the maximum number of generation (i.e. 100) is attained.

■ Feature extraction using features second order statistics:

The second-order statistics depends on the association between two neighbouring pixels in just one offset, where the first pixel is known as the reference and the second one called as the neighbour pixel. The Grey-Level Co-occurrence Matrix (GLCM) is helpful in the characteristic resembling of an image associated with second-order statistics. A GLCM is a matrix where the number of rows and columns is equivalent to the number of grey levels, G , present in the image. The matrix element $P(i, j | \Delta x, \Delta y)$ yields the relative frequency with which two pixels, isolated by a pixel distance $(\Delta x, \Delta y)$, are observed within a certain neighbourhood, one having an intensity 'i' and the other having an intensity 'j'.

The matrix element $P(i, j | d, \theta)$ has the second order statistical probability values for variations between grey levels 'i' and 'j' at a specific displacement distance d and $0^\circ, 90^\circ, 45^\circ$ at certain angle (θ).

In the experiments the GLCM feature descriptor was computed from five values of the angle θ (0 five texture features

(Energy, Contrast Correlation, Homogeneity, Entropy) and displacement distance $d=1$ therefore, the implementation contains 20 features. The texture features are computed as difference Entropy, Difference Variance, Entropy of the Concurrence Matrix, Sum of Squares, Textural Feature, Gradient, Circularity of the Segment, Kurtosis, Skewness, Standard, Mean, Length, Perimeter, Area of the Segmented Mass, Geometric Parameters etc.

▪ **Feature Selection Using Bare-Bones with Adaptive Bat Optimization (BBABO):** The bat algorithm is a novel swarm intelligence optimization technique used for resolving the problems in optimization, and it is influenced by social acts of bats and the phenomenon of echolocation for distance sensing. Along with the cancel of the velocity Bare-bones with Adaptive Bat Optimization (BBABO) does not need any parameter anymore.

A novel Bare Bones -Adaptive Bat Optimization (BBABO) Algorithm with the Gaussian distribution in place of velocity is introduced. In this, the behaviour of the bats is helpful in selecting the significant features from the features extracted for the classification of the healthy and blasted cells.

Echolocation scheme renders the bats the ability of determining the difference between a non-optimal (obstacle) and an optimal features (prey), permitting them to hunt even in complete darkness.

BBABO technique has been designed to act as a band of bats looking out for dimensional reduction $SU = SU = \{f_1, f_2, \dots, f_m\}$ from nucleus and cyclopean image dataset employing their echolocation capability. In the case of BBABO, the echolocation uniqueness is idealized within the constraint of the rules that follow by decreasing the properties by bats:

All of the bats make use of echolocation for sensing the optimal features, and they also are 'aware' of the difference between optimal features (prey/food) and enclosing hurdles in some interesting manner;

Bats fly in random at feature position fp_i with a frequency $freq_{min}$, modifiable wavelength γ and loudness A_0 to achieve increase in the accuracy of tumor segmentation. They can help in repeated fine-tuning of the wavelength of their emitted pulses and they also refine the rate of pulse emission, $r \in [0,1]$ is used.

To begin with, every bat b_i is initialized with starting feature position fp_i , gaussian distribution, and frequency $freq$ are initialized. For every iteration t , and T being the maximum number of iterations, the motion of the virtual bats is decided by updating their velocity and position using equations (8), (9), and (10) as below:

$$freq_i = freq_{min} + (freq_{max} - freq_{min})\beta \quad (8)$$

$$vel_i^t = vel_i^{t-1} + (fp_i^t - fp_*)freq_i \quad (9)$$

$$fp_i^t = fp_i^{t-1} + v_i^t \quad (10)$$

Here, $freq_i$ represents the frequency of every bat, here fp_* stands for the current global optimal error that is found after the comparison of all the solutions obtained among all n number of errors at every iteration ' t '. Once the error position of bats is updated, a random number is generated by using the fuzzy membership function, when the random number is bigger than the pulse emission rate r_i , a new error position will be generated around the current best chosen error, and it can be expressed by equation (11)

$$fp_{new} = fp_{old} + \epsilon A^t \quad (11)$$

Where $\epsilon \in [-1,1]$, refers to a random number, while A^t stands for the average loudness of every bat during the current iteration. Also, the loudness A_i and $f(fp_i) < f(fp_*)$. A_i the pulse emission rate r_i will get updated and a solution will be taken only when a random number is lesser than loudness A_i and $f(fp_i) < f(fp_*)$. A_i and r_i get updated by (12-13)

$$A_i^{t+1} = \alpha A_i^t \quad (12)$$

$$r_i^{t+1} = r_i^o [1 - \exp(-\gamma t)] \quad (13)$$

where α, γ refer to the constants. The algorithm continues iterating until the termination condition is satisfied. In the novel NBOA algorithm, the parameters of the bats like $\beta, \epsilon, A^t, \gamma$ and r_i are fine-tuned by using the Gaussian fuzzy membership function (GFMF). The Gaussian membership function is generally denoted as Gaussian($x; \mu, \sigma$) where μ, σ indicates the mean and standard deviation

$$F_A(x, \mu, \sigma, m) = \exp\left[-\frac{1}{2} \left|\frac{x - \mu}{\sigma}\right|^m\right] \quad (14)$$

Where 'm' is associated with Fuzzification parameter ($m=2$).

ALGORITHM 1. NOVEL BAT OPTIMIZATION ALGORITHM (NBOA):

1. Set the bat population (data samples) fp_i ($i=1, 2, \dots, n$) and vel_i
2. Specify pulse frequency $freq_i$ at fp_i
3. Set pulse rates r_i and the loudness A_i
4. The frequency is adjusted and the velocities are updated to generate new solution. [equations (2) to (4)]
5. Choose a solution among the optimal error
6. Create a local error around the chosen best errors
7. Generate a new solution by flying in random
8. compute A_i via fuzzy membership function
9. Sort the bats and get the current best x_*
10. Post process the results and visualization

This can be combined with efficiency into back propagation during network training

- **Classification Using Fuzzy Neural Network:** Once the essential features are chosen, the classification of healthy and blast cells are performed. Supervised learning in FNN will change its weights of the connections in such a way when the measure of an error is increasingly decreased. Its presentation has to be permissible when the new data is produced.

In this fuzzy neural system, autonomous variables are fuzzified (in case they are crisp variables) and their membership functions are decided at the input unit and suitable rules are made prior their being presented for activating the hidden layer. In addition, the network outputs generated in the output layer are just returned in the form of constants or linear combinations of weighted rules that forms the final output. This generally adds up two units (layers) to the structure.

The process of iteration occurs between the Fuzzification and the defuzzification layers. During this process, suitable weights are estimated for every rule to decide its effect on the output finally.

The input layer of neurons characterizes the input variables to be crisp values. These values are sent to the condition layer that carries out Fuzzification using triangular membership functions with centres denoted to act as the weights into this layer.

A mathematical **system**, which evaluates the analogy input values in terms of logical variables taking on constant values between 0 and 1, in contrary to conventional or digital **logic**, and uses discrete values of either 1 or 0 (true or false, correspondingly).

The outputs produced from the condition layer are sent to the rule layer. This rule layer is similar to a hidden layer of a standard MLP network in terms of its structure and operation. The basic dissimilarity is that in FuNN, every node (neuron) present in the rule layer indicates one fuzzy rule. In the rule layer node the semantic implication is activated. It specifies input data matching with its degree of a corresponding rule. The rule layer produces the output and that output is retrieved by the action element layer. As to the rule layer, the functioning and the structure of the activation layer is similar to the standard hidden layers of MLP networks.

In this layer, every node denotes a fuzzy label from the fuzzy quantisation space of an output variable, for instance, 'small', 'medium' or 'large'. The activation of the node indicates the degree of support which the current data renders to this membership function. The rule layer can be also considered to be an antecedent's layer and the activation layer to be a consequence layer since they signify the antecedent and consequence part of IF-THEN fuzzy rule.

- **Training Algorithm:** A supervised learning system, which carries out categorization, is called as a learner or a classifier. The classifier is given the training data were every data's are already marked with the right label or class. This data is utilized for training the learning algorithm, which generates models and they be utilized for labelling/classifying identical data.

In case there are N samples (x_i, t_i) , where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbb{R}^m$, then the standard SLFN with N hidden neurons and activation function $g(x)$ is computed as:

$$\sum_{i=1}^N \beta_i g(w_i, x_j + b_i) = 0, j=1, \dots, N, \quad (15)$$

Where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ indicates the weight vector, which connects the i th hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ stands for the weight vector, which connects the i th neuron and the output neurons, and b_i stands for the threshold of the i th hidden neuron. The " \cdot " in $w_i \cdot x_j$ specifies the inner product of w_i and x_j . The SLFN attempts to decrease the difference between o_j and t_j .

- **Support Vector Machine:** Support Vector Machines (SVMs) imply supervised learning techniques that are utilized for image classification. It considers the image database given to be two sets of vectors in an ' n ' dimensional space and builds a differentiating hyper plane. This increases the margin between the images having relevance to query and the images with no relevance to the query. There are several pattern matching and machine learning tools and methods used for clustering and classifying the linearly differentiable and non-differentiable data. Support vector machine (SVM) is a considerably novel classifier and its foundation is quite strong arising from the extensive field of statistical learning theory.

A maximum margin classifier forms the fundamental principle of SVMs is. Making use of the kernel technique, the data is inherently mapped onto a high dimensional kernel space. The choice function in the actual space can exhibit non-linearity were the maximum margin classifier have made a firm decision in the kernel space. SVMs categorize the non-linear data present in the feature space into linear data in kernel space.

- **Enhanced Support Vector Machine (ESVM):** Different SVM kernel functions are introduced for users to select for various applications. The most popular kernel functions include the linear function, polynomial function, sigmoid function, and radial basis function. These kernel functions do not take the difference between features of data into consideration. The SVM kernel function format $K(x_i; x)$, shows that every feature of the training or test datasets are considered to be equal. Treatment of all features in the same way may not be effective and it may have an effect on the SVM accuracy. A most probable solution forgiving regard to the significance of various features is to have weights added to a kernel function. The weights are utilized for measuring the significance of every feature. A generic form of the new kernel function is devised to be $K(wx_i; wx)$, where w stands for a vector comprising of weights of features of data set. A nonlinear discriminate function with feature weights is expressed as,

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i y_i K(wx_i, wx) + b) \quad (16)$$

This improved kernel has no dependence on specific kernel functions. As to the diverse applications, one may select the most appropriate kernel function to use the feature weights. Rough set theory is used for calculating and generating these weights from the training data in this research work.

The fundamental principles of weight computation include: 1) when a feature is not present in any reduces, then the weight of this feature becomes 0; 2) the more number of times a feature shows up in the reduces, the more significant is this feature; 3) the lesser the number of features is present in a reduce, the more essential these features showing up in this reduce. In case a reduction contains just one feature, the feature that belongs to this redact is the most vital one.

Once the feature ranking process is completed, those features having 0 weights is considered to be the least significant features and they are deleted.

IV. RESULT AND DISCUSSION

This segment analyse the grades of the experiments carried out on the currentwork. With the help of a tool MATLAB, this work is implemented. In contrast of the existing variableENFBF algorithm and the proposed FNESVM are done in terms of precision, recall and accuracy for the leukaemia image dataset.

PERFORMANCE METRICS

▪ *Precision*: Precision indicates the percentage of the results that show relevance, expressed as

$$\text{Precision} = \frac{\text{True positive}}{\text{true positive} + \text{false positive}} \quad (17)$$

▪ *Recall*: Recall indicates the percentage of the overall relevant results that are categorized by the proposed algorithm that is expressed as

$$\text{Recall} = \frac{\text{True positive}}{\text{true positive} + \text{False Negative}} \quad (18)$$

▪ *Accuracy*: Accuracy is a metric used for the evaluation of the classification models. On an informal note, accuracy defines the fraction of predictions that this model found correctly. Formally, accuracy is defined as:

$$\text{Accuracy} = \frac{\text{True positive} + \text{True Negative}}{\text{Total}} \quad (19)$$

TABLE: 1 performance comparison result values

Methods	Precision	Recall	Accuracy
ENFBF	82.51	6.29	81.33
FNESVM	87.91	8.89	86

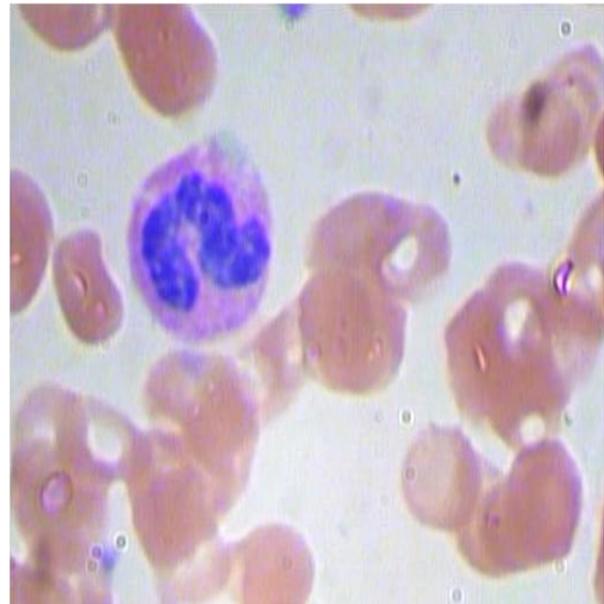


Figure 2: Input image

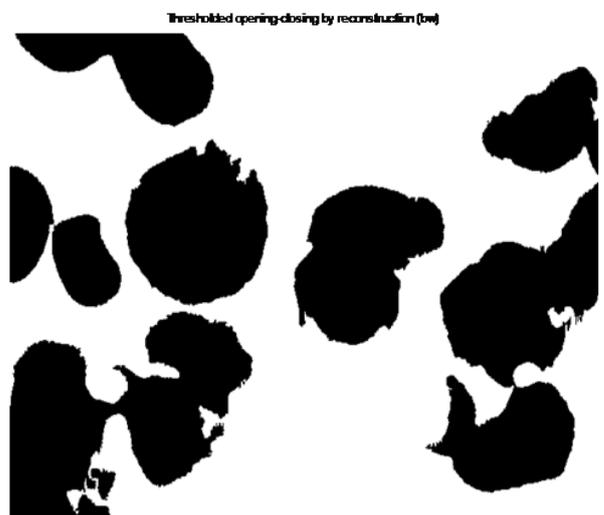


Figure: 3. Threshold opening and closing images

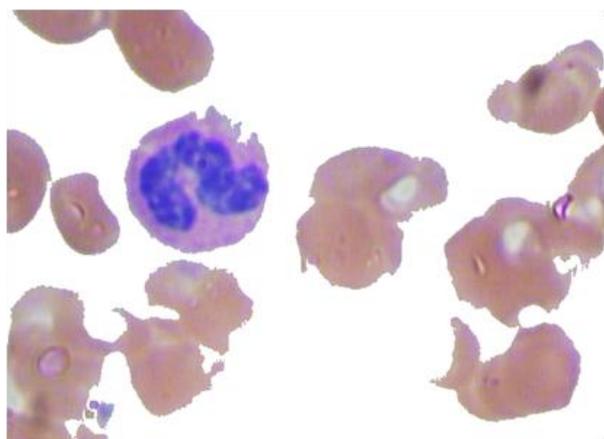


Figure: 4. White blood cells segmented images



Figure: 5. Colored watershed labelled images

Figure.5. illustrates the Coloured labelled images of the input images using Marker Controlled Watershed

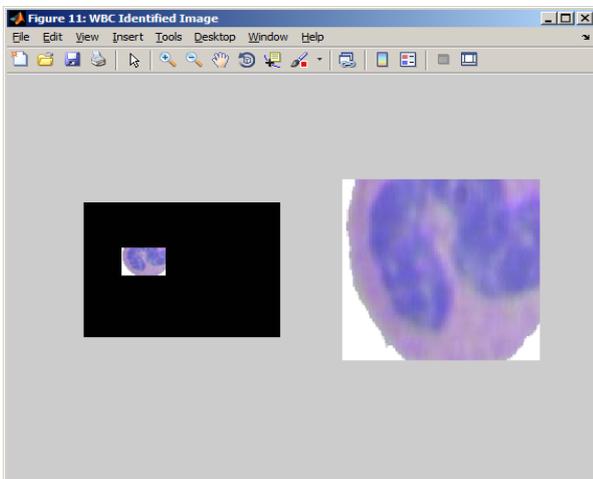


Figure: 6. WBC identified image

Figure 5 illustrates the WBC detected images of the input images using Marker Controlled Watershed segmentation

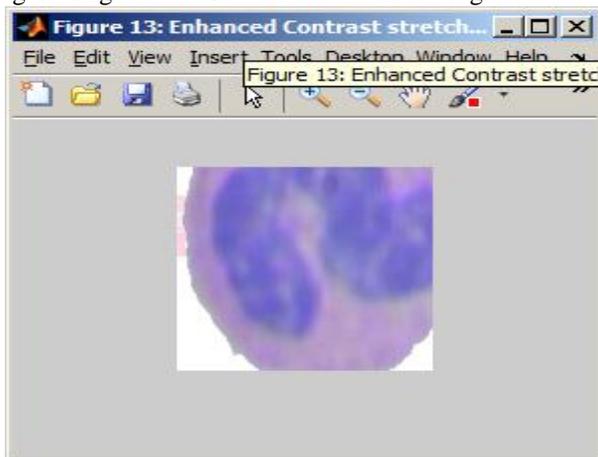


Figure: 7. Image enhancement result

Figure: 6.illustrates the pre-processing of the input image uses Improved Median Filtering for Noise elimination. Figure 7 illustrates the enhancement one of the input image using Enhanced Linear Contrast Stretching.

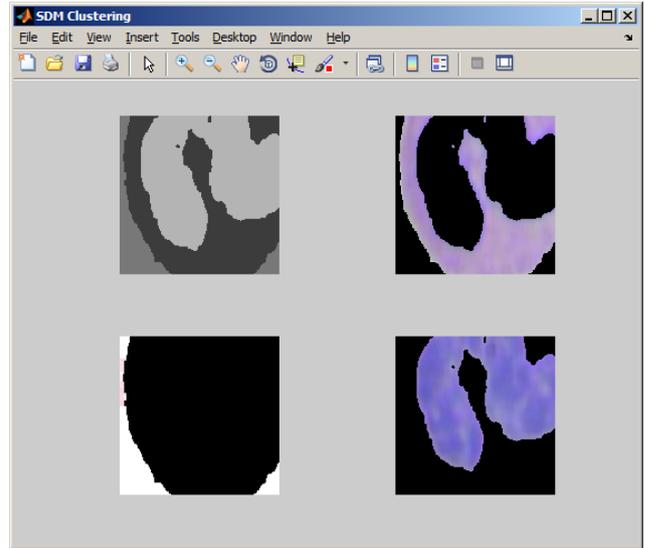


Figure: 8. Segmentation results

Using SDM clustering model, the segmentation is carried out on the pre-processed image for differentiating nucleolus and cytoplasm. The segmented images are illustrated in figure 8.

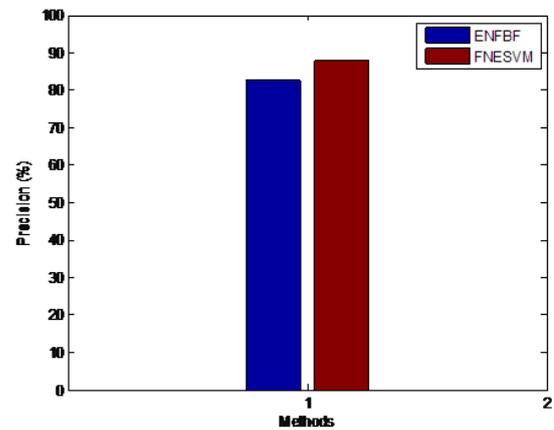


Figure 9. Precision results of different method

Figure 9 illustrates the results of the performance comparison of the proposed FNESVM technique and the already available ENFBF technique in terms of Precision. It can be concluded from the results that the novel model yields much better precision results of 87.91% while the available ENFBF technique yields just 82.51%, correspondingly.

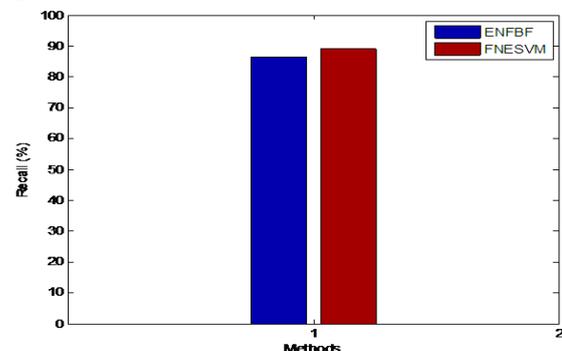


Figure 10. Recall results of different method

Enhanced Support Vector Machine Based Leukaemia Cancer Classification

Figure 10 illustrates the results of the performance comparison carried out between the proposed FNESVM technique and the available ENFBF technique in terms of Recall. It can be concluded from the results that the proposed FNESVM model yields noteworthy Recall results of 88.89 % whereas the available ENFBF technique yields just 86.29%, correspondingly.

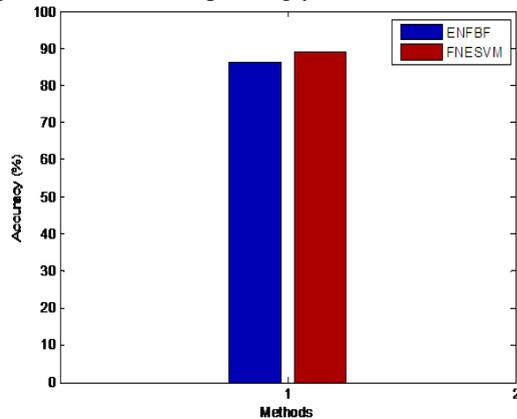


Figure 11. Accuracy results of different method

Figure 11 illustrates the results of the performance comparison carried out between proposed FNESVM technique and the available ENFBF scheme in terms of accuracy. It can be concluded from the results that the proposed FNESVM model yields much better accuracy of 86% while the available ENFBF technique yields just 81.33% correspondingly.

V. CONCLUSION AND FUTURE WORK

In this research work, improved median filtering has been introduced in this technical work, which eliminates the noise from the images by preserving the edges, helping in improving the detection of the leukocytes and lymphocytes during the segmentation process. Enhanced linear contrast stretching is suggested for image enhancement to improve the image. Water marker controlled algorithm is helpful in the WBC identification. SDM is utilized for separating the nucleus and cytoplasm and thereafter feature extraction is carried out by using the second order statistical features. In the proposed technical work, Bare-bones with Adaptive Bat Optimization (BBABO) are introduced for feature selection. At last, classification is carried out by using the Fuzzy Neural Network (FNN) and Enhanced Support Vector Machine (ESVM). As a futuristic approach, it is planned to update the Test data set of SVM to achieve the superior classification accuracy.

REFERENCES

- Shankar, V., Deshpande, M.M., Chaitra, N. and Aditi, S., 2016, October. Automatic detection of acute lymphoblastic leukemia using image processing. In 2016 IEEE International Conference on Advances in Computer Applications (ICACA) (pp. 186-189). IEEE.
- Ching-Hon Pui, M.D, Mary V. Relling, Pharm.D, and James R. Downing, M.D, "Acute Lymphoblastic Leukemia", N Engl J Med 2004; 350:1535-1548 April 8, 2004.
- Mathur A., Tripathi A. S. and Kuse M, "Scalable System for Classification of White Blood Cells from Leishman Stained Blood Stain Images," Journal of Pathology Informatics. HIMA Workshop at MICCAI-2012, Nice, France.
- Joshi, M.D., Karode, A.H. and Suralkar, S.R., 2013. White blood cells segmentation and classification to detect acute leukemia. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2(3), pp.147-151.
- Mohapatra, S., Patra, D. and Satpathy, S., 2014. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Computing and Applications, 24(7-8), pp.1887-1904.
- Patel, N. and Mishra, A., 2015. Automated leukaemia detection using microscopic images. Procedia Computer Science, 58, pp.635-642.
- Harun, N.H., Mashor, M.Y. and Hassan, R., 2011. Automated blasts segmentation techniques based on clustering algorithm for acute leukaemia blood samples. Journal of Advanced Computer Science and Technology Research, 1, pp.96-109.
- Kasmin, F., Prabuwo, A.S. and Abdullah, A., 2012. DETECTION OF LEUKEMIA IN HUMAN BLOOD SAMPLE BASED ON MICROSCOPIC IMAGES: A STUDY. Journal of Theoretical & Applied Information Technology, 46(2).
- Karthikeyan, T. and Poornima, N., 2017. Microscopic image segmentation using fuzzy c means for leukemia diagnosis. Leukemia, 4(1), pp.3136-3142.
- Khobragade, S., Mor, D.D. and Patil, C.Y., 2015, December. Detection of leukemia in microscopic white blood cell images. In 2015 International Conference on Information Processing (ICIP) (pp. 435-440). IEEE.
- Kumar, S., Mishra, S. and Asthana, P., 2018. Automated Detection of Acute Leukemia Using K-mean Clustering Algorithm. In Advances in Computer and Computational Sciences (pp. 655-670). Springer, Singapore.
- Sajjad, M., Khan, S., Shoaib, M., Ali, H., Jan, Z., Muhammad, K. and Mehmood, I., 2016, December. Computer aided system for leukocytes classification and segmentation in blood smear images. In 2016 International Conference on Frontiers of Information Technology (FIT) (pp. 99-104). IEEE.
- Hiremath, P.S., Bannigidad, P. and Geeta, S., 2010. Automated identification and classification of white blood cells (leukocytes) in digital microscopic images. IJCA special issue on "recent trends in image processing and pattern recognition" RTIPPR, pp.59-63.
- Choudhary, R.R., Sharma, S. and Meena, G., 2017, October. Detection of Leukemia in Human Blood Samples through Image Processing. In International Conference on Next Generation Computing Technologies (pp. 824-834). Springer, Singapore.
- Mohapatra, S. and Patra, D., 2010, December. Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images. In 2010 International Conference on Systems in Medicine and Biology (pp. 49-54). IEEE.
- Goutam, D. and Sailaja, S., 2015, March. Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier. In 2015 IEEE International Conference on Engineering and Technology (ICETECH) (pp. 1-5). IEEE.
- Mohapatra, S., Patra, D. and Satpathy, S., 2014. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Computing and Applications, 24(7-8), pp.1887-1904.
- Lei, T., Jia, X., Zhang, Y., He, L., Meng, H. and Nandi, A.K., 2018. Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering. IEEE Transactions on Fuzzy Systems, 26(5), pp.3027-3041.
- Mahabaleshwar, U.S., Sarris, I.E., Hill, A.A., Lorenzini, G. and Pop, I., 2017. An MHD couple stress fluid due to a perforated sheet undergoing linear stretching with heat transfer. International Journal of Heat and Mass Transfer, 105, pp.157-167.
- Xiong, Q., Znamenskiy, P. and Zador, A.M., 2015. Selective corticostriatal plasticity during acquisition of an auditory discrimination task. Nature, 521(7552), p.348.