International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-12, Mar 2019

A Study of Various Algorithms Used for Diagnosing Leukemia From Medical Images

[1]Devika Rajan, [2]Abhay Kolhe,

[1]M.Tech (Scholar), [2]Assistant Professor, Mukesh Patel School of Technology Management and Engineering, Mumbai India.

[1]devika.rajan2003@gmail.com, [2]akkolhe71@gmail.com

Abstract - **Leukemia is a type of cancer that attacks the blood cells which are residing in the bone marrow. It is a cancer that is caused due to increase in the ratio of abnormal white blood corpuscles (WBC) known as blasts in the blood of a human being. The production and functioning of the red blood corpuscles & platelets are severely affected due to the increase in the ratio of these blasts. This can cause loss of life if not identified and cured at an initial stage. Leukemia is diagnosed by manually visualizing and classifying these blood cells from the blood images. This process takes a lot of time and requires manual efforts. The result obtained can be incorrect which results into a delayed diagnosis of the diseases. Image processing is a science that helps us to analyze and modify the Microscopic pictures. Therefore by incorporating various digital image processing techniques a system can be set up for automatic cancer detection. This system takes an input which is usually an image and the output will be the result whether the person is suffering from Leukemia and the type of Leukemia. This system will analyze the blasts or leukocytes present in the blood using image processing techniques and will give the results faster, thus the type of cancer can be diagnosed at an early stage and can be treated accordingly. This paper presents a literature review of various image processing techniques used in various researches for the quick and faultless detection of leukemia.**

***Keywords* — Blasts, Classification, Feature extraction, Pre-processing, Segmentation, White Blood cells.**

I. INTRODUCTION

The blood is produced in a human body through the bone marrow. The stem cell which is a part of the bone marrow produces the main blood cells which are required by our body for its natural functioning. These blood cells are red blood corpuscles (RBC), white blood corpuscles (WBC) and platelets. The WBC's are the guards of our body that fights against all the infection and diseases and prevent their entry into our body. The transportation of oxygen from the lungs to the body's tissues and vice versa is carried out by the Red blood cells or RBC's. Platelets are those blood cells in our body that are capable of forming clots to stop bleeding [1]. A slight variation in the count of these cells is considered as a state of diseases. The normal range of WBC's in our body should be from 3,700 to 10,500. When the count increases beyond this range then the person is said to be suffering from leukemia. The abnormally growing WBC's are known as blasts, they grow and divide at a very fast rate thus crowding and affecting the working of the normal cells. Such cells have a different appearance than the normal cells and they work do not function as expected [2].

There are four different types of leukemia and identifying the type of leukemia will help the doctors to provide a better treatment to the patients. The four types of Leukemia are:

- Acute Lymphoblastic Leukemia
- Chronic Lymphoblastic Leukemia
- Acute Myelogenous Leukemia
- Chronic Myelogenous Leukemia

In acute leukemia the bone marrow cells do not mature at all and they begin to replicate, thus if such leukemia is not identified then the patient only live for few months. On the other hand in chronic leukemia the bone marrow cells do not mature completely as needed so they are unable to fight against the infections like the normal cells. Chronic leukemia patients have longer span of life as compared to acute leukemia affected patients [3].The basic difference between myeloid or lymphocytic leukemia totally depends on the cells in which the cancer develops. The myeloid cells or granulocytes are affected in myeloid leukemia and the lymphocytes are affected in lymphocytic leukemia.

The exact identification of leukemia is done by closely observing and manually analyzing the undeveloped or immature blasts which also includes the lymphoid or

 DOI : 10.18231/2454-9150.2019.0204

myeloid cells. The increase in the range of blasts is also an indication of blood cancer. So the pathologists have to regularly supervise the blood images for the faultless detection of the type of leukemia that the person is suffering. In Spite of the great experience and knowledge of the pathologists this process is time consuming and less accurate [4].Thus there is a need of a system that is capable enough of identifying the leukemia and the type of leukemia without any manual involvement. The main objective of this paper is to present the work and the image processing techniques used by different researchers.

II. LITERATURE REVIEW

A lot of researches have carried out their research in identifying the various image processing techniques that will be able to detect blood cancer and its type from the medical images with better accuracy.

The following set of papers [5], [6], [7], [8], [9], [10], [11] and [12] describes an automatic computerized system that is capable of counting and classifying the WBC'S present in the blood image based on the shape, texture and color features extracted from the infected cells. This automated system is capable of preprocessing the images to remove the noise in the image then extract the Region of Interest (ROI) using any segmentation technique. The image is then represented in terms of some numerical values which are considered to be the feature vector for that particular image and then on the basis of this feature vector the final classification is performed. The feature vector consists of shape, texture and color features. Various techniques of preprocessing has been implemented which includes color conversion, histogram equalization, contrast stretching and so on to enhance the image and to make the next step i.e. segmentation easy. In [7] the cancer detecting system converts the RGB blood image to CIE l_a_b_color format. Channel L holds information about the brightness of the image Channel B holds information about the color features in the blood image which marks the presence of mainly purple, blue shades. Segmentation of the infected cells can be done by retaining back the L and B channel values. This technique helps in precise segmentation of the blasts. Various filters such as median filter, laplacian filter are used for removing the impurities such as noise in the image, which is a part of post processing. Segmentation is a process to isolate the infected WBC cells from other blood components so as to analyze them carefully. K-mean, thresholding, watershed algorithm are some of the most familiar techniques being used. It uses K-mean clustering algorithm for segmentation and for partitioning the overlapping objects watershed algorithm is applied. The combination of these two algorithms provides better segmentation leading to better classification and counting. Feature extraction is the most important step for classification which involves capturing descriptors from the image based on which the experts make their decision. This step extracts three types of feature:

i. Shape: It is numbers which can characterize a given shape of the object (roundness, solidity, area, major and minor axis, perimeter, convex area and perimeter, orientation, elongation, eccentricity)
ii. Color: Theses features are obtained from the color parameters of the image (mean, variance, standard deviation uniformity, smoothness, entropy, energy and skewness)
iii. Texture: It includes the textures properties of the image (energy, entropy, autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, homogeneity)

In [8] classification is described as a process of classifying an unknown data set to a known data set based on some features.SVM classifiers are most commonly used for the classification of the blood sample based on the features extracted. SVM classifier is considered as the best classifier because it does the classification of data set based on hyper plane classifier. An innovative method that segment, classify and count the WBC Blasts simultaneously using a Convolution Neural Network (CNN) is used. The accuracy obtained using a combination of all the three features i.e. shape, color and texture is between 88% - 95%. Some of the limitations are that the sub type is not detected, it does not consider cells in the border of the images and does not consider cell metric images.

Joshi et.al. [13] designed a system that uses Convolutional Neural Network (CNN), Principal Component Analysis (PCA) and a group of 3 classifiers for better classification. The Neural network is trained well to precisely detect and classify the type of leukemia. Segmentation of the image is not included in this approach it directly extracts the features from the image. In this system the segmentation of the cells is not included, instead it uses some pre trained CNNs like CaffeNet, Vgg-f , AlexNet as well as a combination of the three CNNs using the addition of their feature vectors. Features are extracted from the medical images using Principal Component Analysis (PCA). It is the most widely used technique for analyzing data and making predictive models from it. The application of PCA is used for facial recognition identification and object tracking. A combination of three classifiers which includes Support Vector Machine (SVM), Multi-layer Perception and Random Forest are used for the final classification. This provides greater reliability of the results along with accuracy. This system provides complete accuracy (100%) for the detection along with less processing time as compared to other approaches. The future scope of this work is to provide tuning to the architecture in order to improve to get the results in less time with the same accuracy.

 DOI : 10.18231/2454-9150.2019.0204

The following set of papers [14], [15], [16], [17], [18], [19] and [20] describes various image processing steps like image enhancement, segmentation, feature extraction and classification which will executed on Matlab for detection of blood cancer. The feature vectors of these researches consist of only texture features of the image. Preprocessing of the blood image is done to improve the image quality. Various techniques such as color space conversion, histogram equalization, contrast enhancement etc. are done to highlight the area of interest. Various filters like median filter, wiener filter, are used to reduce the unwanted noise present in the image. Gaussian filter helps in smoothening the edges of the blood cells [15]. K-mean segmentation, thresholding, watershed algorithms are some of the techniques done to get the region of interest (ROI) which is the shape and size of the undeveloped or immature WBC cells. An adaptive k-mean clustering is another approach for segmentation in which the user need not specify the number clusters or the initial seed value. Lloyds clustering is also used for partitioning the WBCs from the RBCs and platelets. It helps in finding the centers of the infected blood cells then separated the cells from its background. It gives better accuracy if combined with k-means clustering technique. Watershed algorithm gives an accuracy of 87%, if one is able to differentiate between foreground and background cells which help to detect the boundary [18]. The only limitation of Watershed algorithm is that it often results into over segmentation. Texture features of the cells are extracted using Gray Level Co-occurrence Matrix (GLCM) [14]. GLCM is a famous technique for getting texture feature in an image. One of the basic prerequisite for GLCM is that it can be applied on only on gray scale images. The working of GLCM matrix is done by identify number of times pixel ‘i’ is appearing in vertical, horizontal or diagonal position to a pixel ‘j’, where ‘i’ and ‘j’ are gray scale intensity values of the cells. The Gray Level Difference Method (GLDM) also helps in texture feature extraction. It is difference in the intensity values of two gray pixels. Some of the texture features extracted are energy, entropy, correlation, variance homogeneity, cluster shade and dissimilarity. Homogeneity of an image is the degree of variation for the texture values. The value of homogeneity is more if the variations are less between the intensity values. The range of homogeneity is between 0 and 1.The relation between the neighbour pixels values is represented by correlation. Texture feature extraction extracts values that are non-redundant and informative that helps in better detection of the disease. The features along with SVM classifier help in the precise detection of the type of leukemia the person is suffering. The accuracy obtained when the feature vector contains only texture features is between 85% - 92%.

Madhloom et.al. [21] and [22] focuses on automatic detection of acute lymphoblastic leukemia in their researches using the shape and geometric features. These methods have a scope for early detection of the disease with high speed and accuracy. The preprocessing of both the work includes color space conversion because the colors in the RGB color space are very much close to each other which make the segmentation difficult. Thus in this system segmentation is done by converting the RGB image to CIE L*a*b color space in which the pixel value of each object is distinguishable. Segmentation can also be done by converting the blood image from RGB plane to CMYK color space then thresholding is applied to discard the background and retain the region of interest. Since the Lymphocytes are immature WBC they will have a different color as compared to the mature WBC's so in the CMYK color space these cells will have a high contrast color. After the conversion the WBCs are extracted by only retaining only the pixel values that hold the blasts. Both the work does not include any other segmentation technique. Shape features like ratio of nucleus to cytoplasm are calculated from the segmented image. The nucleus of the cell is said to have brighter pixels as compared to the cytoplasm. It also takes in consideration some more shape features like Sate of chromatin, Number of holes in Nucleus and cytoplasm. A fuzzy set of rules are formulated for these 4 parameters which decide whether the person is leukemic or not. Shape features such as roundness, perimeter etc. are extracted from the cells. Final classification is done on the basis of the count and the features extracted from the segmented WBC's. One of the main advantages of this technique is that they are able to detect the severity of Acute Lymphoblastic Leukemia (ALL) Cancer. The accuracy rate achieved using these techniques are more than 90%. The main benefit of this technique is that it has a self-learning mechanism that helps in better detection of the cancer. The accuracy can be improved by incorporating some more segmentation techniques with the existing one’s which would give an exact segmentation of the infected cells leading to better detection.

The following set of papers [23], [24], [25], [26], [27], [28], [29], [30] and [31] presents an innovative method for cancer detection in which the feature vector consist of texture and geometric parameters of the image. It uses Contrast enhancement for improving the image quality as a part of preprocessing. Color conversion is also another important and best technique for preprocessing because it highlights the infected cells with respect to its background. The input image is converted from RGB to L*a*b color space. All the color information of the input image is thus represented in a* and b* which helps in nucleus segmentation. Segmentation plays a very important role for the preceding steps i.e. feature extraction and classification. K-mean is considered as good segmentation technique for an image of any size and quality. The Number of Clusters in K-mean is based on user’s choice. Ideally the image can be

 DOI : 10.18231/2454-9150.2019.0204

divided into 4 clusters Nucleus, cytoplasm, background and other cells. The segmentation can also be done using Zack algorithm in combination with K-mean algorithm. This algorithm helps in detection of the myeloid cells. Roundness of the cells is calculated to identify the single and grouped cells. The value of roundness indicates whether the cell is circular or not. It is calculated by dividing the area by perimeter. If the value is 1 then it indicates that the cell is circular else it is not. A threshold value is selected and all the cells having value less than the threshold is said to be grouped cells and the ones having value greater than threshold is single cell. Post processing is done to avoid the results of over segmentation. Feature extraction is a process of converting the images details into numerical values that can be used for classification. Shape and texture features of the image are extracted to detect the cancer. Shape features describe the shape of the object. The object includes nucleus cytoplasm. Texture features describes the texture parameters and the variation in the pixel intensities. Some of the shape features are area, perimeter, roundness, major axis, minor axis, elongation, eccentricity etc. The texture features such as energy, entropy, contrast, correlation, homogeneity are used. Ritter and Cooper in their study use a combination of dual neural network having a Fuzzy ARTMAP and simplified Fuzzy ARTMAP Neural Network for the precise detection [28]. This technique can only identify if the person is suffering from leukemia or not, it does not predict the type of leukemia. The advantage of this approach is that it gives a higher rate of accuracy with the help of a dual neural network. SVM classifier is being used by rest of the researchers because of the reliable results given by the network. The accuracy obtained by these techniques is about 85% to 94%. The limitations in [25] and [29] research is that the techniques do not detect the sub types and severity of the cancer. Better segmentation and feature extraction of the nucleus and cytoplasm will also help to improve the accuracy.

In the following set of papers [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43] and [44] discusses a set of image processing techniques that are used for locating and counting the undeveloped WBC's present in the biomedical images by extracting only the shape features of the blasts. Preprocessing is the step for analyzing the medical images which involves cleaning the image and removing the noise from the image and improving the image quality. Preprocessing and segmentation is applied to both the training and testing images. The Preprocessing mainly includes conversion of blood image from RGB to different color spaces which focus on the Region of Interest. The RGB image is converted in gray scale image and then the contrast is enhanced to highlight the infected WBC's. Converts the image into YCbCr space where Y stands for Luminance, Cb stands for Blue Value, Cr stands for Red Value. After the conversion the Cb and Cr coefficients are extracted. Rangolein et al. [32] in their study converts the image into binary image by selecting an appropriate threshold value which will have the WBC's in white pixel and the background and RBC's in black pixel value. A set of filters such as median filter, laplacian filters are used for noise removal, preserving the edges and to darken the nuclei of the cells. In [35] it uses Watershed algorithm which is a simple and fast computation method for segmentation but cannot be applied on every image, it givens an accuracy of 72%. K-mean segmentation though it is a good clustering technique and it have an accuracy rate of 72.3% [40]. The other techniques like histogram equalization and linear contrast stretching method that can detect the white blood cells but they cannot differentiate between overlapping cells thus, it gives an accuracy of 73%. A combination of two or more segmentation techniques can yield better result. In the post processing stage various morphological operations like dilation and opening is performed to have well-defined boundaries for the cell. Other morphological operations such as hole filing and erosion are also applied to locate the exact boundary of the WBC cells. The detection is done by identifying the shape based features they work well with overlapping objects but give an approximate result. It has an accuracy rate of 95%. Geometric features such roundness, solidity, area, major and minor axis, perimeter, convex area and perimeter, orientation, elongation, eccentricity, rectangularity, compactness, convexity are calculated for better detection of the disease. The count number of pixel inside nucleus of blasts then the area of WBC cells can also be used to identify the severity of the diseases. These researches only focus on the shape and geometric features of the nucleus and cytoplasm for the detection. It is found that the shape features of the nucleus hold more information required for the detection as compared to the shape features of the cytoplasm. The rate of accuracy is greater when Shape features are part of the feature vector more than 90% accuracy is obtained in each case.

In the following set of papers [46], [47], [48], [49], [50] and [51] it presents a system for identifying acute and chronic leukemia from a patient's blood image by extracting the texture features using Local Binary pattern. Acute leukemia grows faster as compared to the chronic leukemia. Preprocessing in these researches use thresholding and contrast stretching to improve the quality of the blood smear. Gabor filter is a type of linear filter that is used for analyzing the texture of the image. It is a filter which has application in pattern recognition. The Segmentation includes morphological operations along with discrete Fourier transform algorithm to separate the overlapping objects or cells. K-mean clustering algorithm is also used since it is considered to be one of the best segmentation techniques because it is versatile and can be used segmenting any type of data. K-mean clustering is

 DOI : 10.18231/2454-9150.2019.0204

performed for segmenting and retaining the undeveloped WBC's from the RBC and platelets. The feature vector of these papers includes Texture features that are extracted from the Local Binary Pattern (LBP) Image of the blood smear. LBP is a very famous technique for extraction of texture features required for pattern recognition [48].

In the LBP image every pixel values is computed based on its immediate 8 neighbour's. The first step is converting the RGB Image into gray image. It then uses a 3*3 mask to calculate the LBP pixel values. The pixels among the 8 neighbour having value greater than the central pixel is replaced by 1 and the pixels having value less than or equal to central pixel is replaced by 0. After that the new decimal value is calculated from its binary representation using a formula. In this way every pixel of original image is replaced by its LBP Value. The LBP Image obtained holds more of texture parameters of the image. An example of LBP pixel value generation is as shown below:

Figure 1: First step in LBP Creation

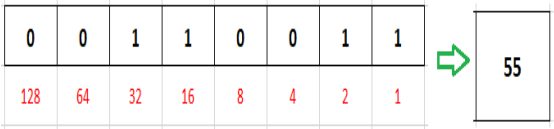

Figure 2: Calculation of the new pixel value

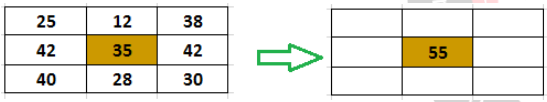

Figure 3: Replacing original value with the new value

After getting the LBP Image the texture features are then calculated from it. These texture features include energy, entropy, auto-correlation, contrast, correlation, homogeneity etc. These feature vectors are then passed as input to the Support Vector Machine (SVM) classifier. The SVM classifier along with Artificial Neural Network is used for accurate classification. Such a combination reduces the time of processing and increases the efficiency and accuracy of the network to more than 90%. [49] It uses a sequential two step neural network classifier that improves the result of detection. The accuracy obtained using these techniques are between 84% to 92%.The accuracy can be improved by including some more features along with the texture features.

The following set of papers [52], [53], [54] and [55] describes about detection of leukemia by only counting the number of WBC in the image and extracting some geometric features from it. Segmentation of the digital image is done by converting the image from the existing color space to CIE l*a*b color space. This method purely identifies the sub categories of acute leukemia. The Segmentation is followed by morphological operations along with Fourier Transform method to separate the overlapping objects or cells. Geometric and color features are extracted from the segmented objects. Otsu's threshold method is also used for enhancing the image and partitioning the affected cells from the normal RBC and platelets. After the exact segmentation, required features are taken from the images and given as an input for the classification process. The pre-processing of the input image is done by contrast enhancement and various morphological operations. PCA algorithm is used for feature extraction that extracts main features namely shape feature. These features help to differentiate to between normal and abnormal cells. The total count of infected cells is also considered as an important parameter for the detection. For better detection of the cells edge detectors like canny and sobel can be used that provides precise edge detection of the cells. The two step neural network classifier provides an accuracy of 97.7%. K Nearest- Neighbour (kNN) Classifier is used to label the affected cells and the patient as leukemic or non-leukemic. The computation in kNN classifier is very simple since it compares the input data with the data present in its knowledge base and finds a majority match for that input data and classifies that data into that class. It can also be used to differentiate blast cells from unaffected white blood cells. The kNN classifier requires a lot of training data for a better prediction. The only disadvantage of kNN Classifier is it will not give expected results when the image contains noise [53]. It gives an accuracy of 93%.

The literature review of all these papers can be consolidated as shown in the table 1.

Table 1: Consolidated analysis of literature review

	Classification Parameters				
Citation	**Shape**	**Color**	**Texture**	**Limitations**	**Accuracy**
[5]-[12]	✓	✓	✓	It does not detect the sub types of leukemia, it does not consider cells in the border of the images and does not consider cell metric images.	88% - 95%

 DOI : 10.18231/2454-9150.2019.0204

[13]	PCA			This system is not able to perform the task on larger image set therefore the future scope of this work is to provide tuning to the architecture in order to improve to get the results in less time with the same accuracy.	100%
[14]–[20]			✓	The future work is to improve the accuracy by improving the techniques used in the system. The future work includes in reducing the processing time by identifying alternate techniques to extract the texture features.	85% - 92%.
[23]-[31]	✓		✓	The limitations in [25] and [29] research is that the techniques do not detect the sub types and severity of the cancer. Better segmentation and feature extraction of the nucleus and cytoplasm will also help to improve the accuracy	85% to 94%.
[32]-[44]	✓			The future work is to improve the segmentation technique to yield better results of detection. A combination of two or more segmentation algorithms will help in precise detection	More than 90%
[46]-[51]	Texture features using Local Binary pattern			The accuracy can be improved by including some more features along with the texture features.	84% to 92%.
[52]-[55]	Count of WBC's present			The future work is to improve the classification and detect the subtypes of leukemia by analyzing the shape of the cells	90% -93%

III. RESEARCH METHODOLOGY

From the literature review it is clear that for the precise detection of leukemia and the type of leukemia the following steps have to be incorporated in the automatic cancer detecting system. The system will have two main phases: the initial phase is known as the training phase and the next phase is the testing phase.

Training Phase: In the Training phase the classifier is trained with a set of images whose specified class is already known to the user. In this case the system is trained with set of microscopic images which are previously classified as cancerous or non-cancerous. All the required features will be extracted from these images. These features along with their classification will be stored in the knowledge base for future classification or prediction. The success or accuracy obtained in the testing phase depends on the training and the data present in the knowledge base.

Testing phase: The steps performed in testing and training phase are almost the same but, the only and major difference is that the input query given in this phase is not classified and the system has to predict whether the input belongs to the cancerous group or non-cancerous group. The input image is processed first and then segmented to retain the region of interest. After getting the region of interest all the features that were extracted in the training phases are extracted from this image and on the basis of these features and the data present in the knowledge base the system predicts whether the person is leukemic or not.

The flow of activities in the system is as shown in figure 4.

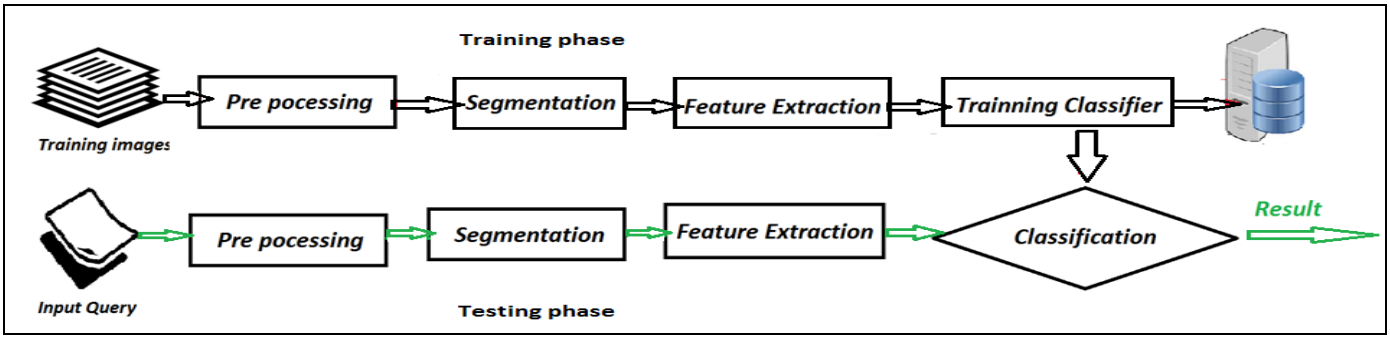

Figure 4: Basic flow of activities

1. Image Acquisition: The biomedical images required for both the testing and training phases can be acquired from any pathological labs, medical laboratory or we can also get images from online database that are available for free of cost. The Acute Lymphoblastic

 DOI : 10.18231/2454-9150.2019.0204

Leukemia Image Database also known as ALL-IDB consists of images of cancerous and healthy lymphocytic cells. All the images in this database are uniform in nature and can be used for various image processing activities like segmentation feature extraction, classification etc

2. **Image Preprocessing:** Pre processing of an image is done with an intension of enhancing the given image so that it yields better results in the proceeding steps. Noise in an image is a component that will unintentionally affect the result obtained at every step thus, having a noise free image is very essential for the precise and accurate detection of the disease. In preprocessing step various digital image processing techniques are performed on the image to make it noise free and to enhance the quality of the image. The quality of the image is in terms of its brightness, contrast and sharpness. The edges of the blasts have to be clear for proper segmentation and feature extraction. Thus this step focuses on denoising and improving the quality and visual appearance of the image. Some the techniques used are color conversion, contrast enhancement, histogram equalization filtering etc.

In color conversion the RGB image is converted into another color space like CMYK color space, CIE L*a*b color space etc. The WBC's have a bright color as compared to the RBC's and the platelets therefore a simple color conversion will highlight the color of the blasts. Contrast enhancement and histogram equalization is done in order to improve the color parameters of the image. The goal is to highlight the WBC's present in the image so that they can be easily segmented from the other components. Filters are used to discard the unwanted components such as noise from the image.

3. **Image Segmentation:** The blood images acquired from the laboratory will have WBC, RBC and platelets in it. The area of interest for blood cancer detection is only the immature WBC rest of the portion has to be discarded. Image segmentation is the technique which is used for this activity. It will separate the foreground objects from its background using different techniques. There are lots of techniques available but a single segmentation algorithm will not be applicable for all the images. The quality of the segments obtained at the end using a single segmentation algorithm will be different for different images because some of the images may have overlapping cells and some may not have them. Therefore researchers implement a combination of two or more segmentation technique to get better segments. Sometimes over segmentation causes modification to the shape of the cell which will lead to a wrong detection, so to avoid this various morphological operations can be applied to regain the shape of the segmented blasts. K-mean segmentation, watershed algorithm, Fuzzy c means algorithm and thresholding are some of the most commonly used segmentation algorithms.

Thresholding is the simplest segmentation technique where in the gray image or a color image is converted into a binary image based on the threshold value [32]. All the pixels having a value greater than the threshold value is set to 1 and all other pixels are set to 0. Since the white blood cells are the brighter cells a simple thresholding will help in extraction of the blasts. The threshold value can be selected using the local thresholding method where the mean value of all the pixels in the image is computed and that will serve as the threshold for that image [10].

Kmean segmentation technique is a famous clustering technique [12]. Initially it begins by selecting k cluster centers randomly.

1. It will then compute the Euclidian distance of every pixel with each cluster center. This pixel will then belong to the nearest cluster based on the Euclidian distance.
2. In the next step the cluster centers are calculated by taking the mean of all the pixels present in every cluster.
3. The algorithm in the next iteration will compute the Euclidian distance of each pixel with the new cluster center values obtained and group the pixels in the same way.
4. The cluster centers are calculated again by taking the mean of all the pixels belonging to that group.

These steps are repeated until the mean value of each cluster remains the same

Watershed segmentation technique is another classic technique used for segmentation of objects from the background. In this technique the algorithm first calculates the distance transform of the image. This is a matrix that represents the distance of every pixel from its nearest 0 pixel value [18]. This distance transform matrix will help us to extract the boundary of the objects. These boundary lines are called as the watershed lines. The pixels belonging to the either sides of this line are said to be part of the catchment basin. Using the watershed lines we extract the objects from its background.

4. **Feature extraction:** Feature extraction plays an important role in cancer detection. In this step all the features required for deciding whether the person is suffering from leukemia are extracted from the segmented blasts that are obtained in the previous step. The number of features extracted in the testing phase

 DOI : 10.18231/2454-9150.2019.0204

and training phase will be the same. In the training phase the feature details are stored into the knowledge base for future reference. These features include shape, color, location, count, geometric and texture.

- **Color** - It has been observed that the color of immature or undeveloped WBC have a darker shade than the RBC and the platelets so, the color feature can be extracted as one of the parameter for the counting and identifying the type of leukemia.
- **Shape** - The shape features include the roundness eccentricity of the blasts. In Acute Lymphoblastic or Chronic Lymphoblastic Leukemia the shape of the nucleus is round where as in Acute Myeloid or Chronic Myeloid Leukemia the shape of the nucleus is curved.
- **Size** - The Size features include calculating the area, perimeter solidity etc. of the white blood cells. It is observed that the size of lymphoblast's are less than the size of the other blasts
- **Texture** - The homogeneity and energy are the main texture features that are extracted from the WBC's some other texture features include auto correlation, contrast etc.
- **Geometric** - The elongation of the nucleus, compactness, concavity, radius etc. are the geometric features that are extracted from the affected blood cells

5. Image Classification: It is the final and main step for the detection of the cancer. The success of this stage completely depends on the training given to the classifier. In the training phase the classifier is trained with set of images which are already classified as cancerous or non-cancerous. This classification along with their features is stored in the knowledge base. When a new unclassified input arrives the classifier will extract the features from that input and then compare them with the features present in the knowledge base. When an approximate or exact match is found with the feature values present in the knowledge base then the classifier will classify the new input into that class which is specified in the knowledge base. The different sub categories of leukemia can also be specified in the training phase so that the new input can also be classified into its subtype. Some of the classifiers used for classification are SVM classifier, neural networks etc. SVM classifier is considered as the best classifier because it does the classification of data set based on hyper plane classifier. The input set of data is separated into two classes in such a way that it optimizes the margin between the two classes. During the training phases features obtained are said to be the support vectors, based on these support vectors the unknown data is classified [23].

IV. Conclusion & Futurework

When the count and size of the white blood cells increases with respect to the other blood cells then the person is said to be leukemic. This work presents a study of various methods developed by different researchers for the accurate diagnosis of Leukemia cancer from the blood images. The series of activities performed for the detection includes image enhancement for improving the contents present in the image. Color conversion, histogram equalization, contrast stretching are some of the pre processing techniques which are used by different researchers. It is then followed by segmentation of the affected white blood cells for calculating the feature values. A proper segmentation of the cells will yield better detection of the disease. The feature values are calculated from the segmented image which includes the count of white blood cells, shape, color and texture properties. Therefore it is very essential to have segmentation technique that will precisely extract only the white blood cells from the blood image. The feature values along with the data in the knowledge base will provide the final classification value. In order to have a better accuracy we have to train the classifier with more number of images. Accuracy of the system increases with increase in the training given to the system. Thus we have to consider all these parameters before designing a cancer detecting system.

The future work proposed is to design a cancer detecting system by pipelining various image processing techniques that are discussed in the literature review. This system must be capable of delivering promising results with less computational efforts and processing time. This system should also be able to detect the subtypes of leukemia along with its severity. An early and precise detection of the leukemia and its type will help the doctor to provide proper medication before it's too late. An early detection will not only save a life but also reduce the cost required for treatment considerably

References

[1] C.R., Valencio, et.al. ,"Knowledge Extraction Using Visualization of Hemoglobin Parameters to Identify Thalassemia",Proceedings of the 17th IEEE Symposium on Computer Based Medical Systems, 2004, pp.1-6.

[2] Adollah et.al. , "Blood Cell Image Segmentation: A Review", Biomed2008, Proceedings 21, 2008, pp. 141-144.

[3] M.Saritha, Sukesh.K, Prakash B.B, Srinivas. B, "Detection of Blood Cancer in Microscopic Images of Human Blood Samples: A Review" in International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), March 2016

[4] A. R. Jasmine Begum and T. Abdul Razak," DIAGNOSING LEUKEMIA FROM MICROSCOPIC IMAGES USING IMAGE

 DOI : 10.18231/2454-9150.2019.0204

ANALYSIS AND PROCESSING TECHNIQUES",2017 World Congress on Computing and Communication Technologies,2017

[5] Basima C.T. and Janu R Panicker "Enhanced Leucocyte Classification for Leukaemia Detection", 2016 International Conference on Information Science (ICIS), pp 65 - 71, 2016

[6] Luis Henrique Silva Vogadoet.al."Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-trained Convolutional Neural Networks", Graphics, Patterns and Images (SIBGRAPI), 2017 30th SIBGRAPI Conference, 2017

[7] Ashwini Rejintal and N. Aswini,"Image Processing Based Leukemia Cancer Cell Detection", IEEE International Conference,pp 471-474 , Jan 2017.

[8] Mohammed Bilal N Shaikh and SachinDeshpande,"Computer Aided Leukemia Detection using Digital Image Processing Techniques",IEEE International Conference,pp 344-348, 2017

[9] ,R. Hassan et.al, "Leukaemia Screening Based on Fuzzy ARTMAP and Simplified Fuzzy ARTMAP Neural", 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences,pp 11-16, 2012

[10] Ivan Vincent et.al. ,"Acute Lymphoid Leukemia Classification using Two-Step Neural Network Classifier", Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop, pp 1-4, 2015

[11] Vasuki Shankar et.al. ,"Automatic detection of Acute Lymphoblastic Leukemia using Image Processing", 2016 IEEE International Conference, pp 186-189, 2016

[12] Sachin Deshpande et.al. ,"Leukemia Detection using Digital Image Processing Techniques", 2017 IEEE International Conference, pp 222-229,2017

[13] Ms. Minal D. Joshi, Prof. Atul H. Karode, Prof. S.R.Suralkar, "White Blood Cells Segmentation and Classification to Detect Acute Leukemia" in International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 3, May – June 2013

[14] N., H., A., Halim, M., Y., Mashor, R., Hassan, "Automatic Blasts Counting for Acute Leukemia Based on Blood Samples", International Journal of Research and Reviews in Computer Science, Vol. 2, No. 4, August 2011, pp. 971-976.

[15] NiranjanChatap et.al. ,"Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbor", eISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III (Sep – Oct. 2014), PP 79-87 www.iosrjournals.org..

[16] A.Sindhu1, S.Meera2 " A Survey on Detecting Brain Tumor mri Images Using Image Processing Techniques" , International Journal of Innovative Research in Computer and Communication Engineering. Vol. 3, Issue 1, January 2015.

[17] Scotti, F., "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," Computational Intelligence for Measurement Systems and Applications, 2005. CIMSA. 2005 IEEE International Conference on , pp.96,101, 2005.

[18] Mohapatra, S.; Patra, D.; Satpathi, S., "Image analysis of blood micro- scopic images for acute leukemia detection," Industrial Electronics, Con- trol& Robotics (IECR), 2010 International Conference on , pp.215,219,2010.

[19] Mohapatra, S.; Patra, D.; Satpathy, S., "An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images," Neural Computing and Application, 2013. Springer-Verlag,pp.1,18, June 2013.

[20] Nor Hazlyna, H.; Mashor, M.Y.; Mokhtar, N. R.; Aimi Salihah, A.N.;Hassan, R.; Raof, R. A A; Osman, M.K., "Comparison of acute leukemia Image segmentation using HSI and RGB color space," Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on , pp.749,752, 2010.

[21] Madhloom, H.T.; Kareem, S.A.; Ariffin, H., "A Robust Feature Ex-traction and Selection Method for the Recognition of Lymphocytes versus Acute Lymphoblastic Leukemia," Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conferenceon , pp.330,335, 2012.

[22] S.Jagadeesh ,Dr.E.Nagabhooshanam , Dr.S.Venkatchalam , "Image Processing Based Approach to detect Acute Lymphoblastic Leukemia in Blood Sample",International Journal of Technology and Engineering Sciences, Vol.1 (1), ISSN:2320-8007.

[23] S.Jagadeesh ,Dr.E.Nagabhooshanam , Dr.S.Venkatchalam , "Image Processing Based Approach to Cancer Cell Prediction in Blood Sample",International Journal of Technology and Engineering Sciences, Vol.1 (1), ISSN:2320-8007.

[24] N., Ritter, J., Cooper, "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", 30thAustralasian Computer Science Conference, Conference in Research and Practice in Information Technology, Vol. 62, 2007, pp. 161-169.

[25] Abdul Nasir, Mustafa N, Mohd Nasir, " Application of Thresholding in Determing Ratio of Blood Cells for Leukemia Detection" in the Proceedings of International Conference on Man-Machine system (ICoMMS) oct 2009 ,BatuFerringhi, Penang, Malaysia

[26] L. Putzu, C. Di Ruberto, "White Blood Cells Identification and Classification from Leukemic Blood Image", Proceedings of the IWBBIO , 2013

[27] Mohapatra S, Patra D. Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images. 2010 IEEE International Conference on Systems in Medicine and Biology (ICSMB)2010:49–54.

[28] N., Ritter, J., Cooper, "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", 30thAustralasian Computer Science Conference, Conference in Research and Practice in Information Technology, Vol. 62, 2007, pp. 161-169.

[29] D.M.U., Sabino, L.D.F., Costa, L.D.F., E.G., Rizzatti, M.A., Zago, "A Texture Approach to Leukocyte Recognition", Real Time Imaging, Vol. 10, 2004, pp. 205-206.

[30] Dorini, L.B, Minetto, R, and Leite, N.J, 2007 'White blood cell segmentation using morphological operators and scale-space Analysis' Unicamp - UniversidadeEstadual de Campinas Instituto

 DOI : 10.18231/2454-9150.2019.0204

de Computac¸ ˜aoCaixa Postal 6176, 13083-971, Campinas, SP, Brasil.

[31] Devesh D, Nawgaje, And Kanphade, R. D, 2011 'Implementation of Fuzzy Inference System For White Blood Cell Cancer Detection Using DSP TMS320C6711', International Journal of Engineering Science and Technology (IJEST) ISSN : 0975-5462.

[32] Mohamed, M.; Far, B.; Guaily, A., "An efficient technique for white blood cells nuclei automatic segmentation," Systems, Man, and Cyber- netics (SMC), 2012 IEEE International Conference on , pp.220,225, 2012.

[33] NiranjanChatap and Sini Shibu ―Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbor‖, e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III (Sep – Oct. 2014), PP 79-87.

[34] BakhtAzam, Rashid Jalal Qureshi, Zahoor Jan, and Taj Ali Khattak ―Color Based Segmentation of White Blood Cells in Blood Photomicrographs Using Image Quantization‖, Research Journal of Recent Sciences, Vol. 3(4), 34-39, April (2014).

[35] Nagabhushana R M and Rajeshwari P ―Geometrical Analysis of Leukocyte Nucleus to Detect Lymphoblast from Microscopic Blood Images‖, International Journal of Technology and Engineering Science [IJTES]TM Volume 2[6], pp: 1947-1951, June 2014.

[36] C.R., Valencio, et.al. ,"Knowledge Extraction Using Visualization of Hemoglobin Parameters to Identify Thalassemia",Proceedings of the 17th IEEE Symposium on Computer Based Medical Systems, 2004, pp.1-6.

[37] Adollah et.al. , "Blood Cell Image Segmentation: A Review", Biomed2008, Proceedings 21, 2008, pp. 141-144.

[38] M.Saritha, Sukesh.K, Prakash B.B, Srinivas. B, "Detection of Blood Cancer in Microscopic Images of Human Blood Samples: A Review" in International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), March 2016

[39] C. Raje and J. Rangole, "Detection of leukemia in microscopic images using image processing," International conference on communication and signal processing, April 2014.

[40] F. Kashmin, A. S. Prabuwono, and A. Abdullah, "Detection of leukemia in human blood sample based on microscopic images: A study," Journal of theoretical and applied information technology, December 2012.

[41] S. Rajendran, "Image acquisition and retrieval systems for leukaemia cells," in Mastersthesis, Department of Computer Science, UM, 2007.

[42] D. Joshi, M, H. Karode, A, and R. Suralkar, S, "White blood cells segmentation and classification to detect acute leukemia," in International journal of emerging trends and technology in computer science, June 2013.

[43] G. Zadeh, H, S. Janianpour, and Haddadnia, "Recognition and classification of the cancer cells by using image processing and labview," International journal of Theory and Engineering, vol. 5, February 2013

[44] J. N. Jameson, L. K. Dennis, T. R. Harrison, E. Braunwald, A. S. Fauci, S. L. Hauser, and D. L. Longo, "Harrisons principle of internal medicine. newyork," 2005.

[45] F. Kasmin, A. S. Prabuwono, and A. Abdulah, "Detection of leukemia in human blood sample based on microscopic images a study," in Journal of Theoretical and Applied Information Technology, vol. 46, December 2012.

[46] Nanni, L.; Lumini, A.; Brahnam, S., "Local binary patterns variants as texture descriptors for medical image analysis," US National Library of Medicine National Institutes of Health , pp.117,125, 2010.

[47]RobertM.Haralick,KShanmugam,ItsakDinstein"Textural features for image classification."IEEETransaction On system man and cybernetics (610- 621) volume3-No.6, November 1973.

[48] N. Mustafa, N. A. MatIsa, M. Y. Mashor, and N. H. Othman, "Colour contrast enhancement on preselected cervical cell for thinprep images," in Proceedings of the ThirdInternationalConferenceonInternationalInformation Hiding and Multimedia Signal Processing (IIH-MSP 2007), Kaoshing, Taiwan, 2007.

[49] M. Wang, X. Zhou, F. Li, J. Huckins, R. W. King, and T. C. Wong, "Novel cell segmentation and online learning algorithms for cell phase identification in automated time-lapse microscopy," in Biomedical Imaging: From Nano to Macro, ISBI 2007, pp. 65–68, 4th IEEE International Symposium, April 2007.

[50] Leonardo de Oliveira Martins, Geraldo Braz Junior, Arist´ofanes Correa Silva, Anselmo Cardoso de Paiva, and Marcelo Gattass (2009) "Detection of Masses in Digital Mammograms using K-means and Support Vector Machine" Electronic Letters on Computer Vision and Image Analysis,vol 8(2), pp.39-50.

[51] N., Ritter, J., Cooper, "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", 30thAustralasian Computer Science Conference, Conference in Research and Practice in Information Technology, Vol. 62, 2007, pp. 161-169.

[52] Abdul Nasir, Mustafa N, Mohd Nasir, " Application of Thresholding in Determing Ratio of Blood Cells for Leukemia Detection" in the Proceedings of International Conference on Man-Machine system (ICoMMS) oct 2009 ,BatuFerringhi, Penang, Malaysia

[53] L. Putzu, C. Di Ruberto, "White Blood Cells Identification and Classification from Leukemic Blood Image", Proceedings of the IWBBIO , 2013

[54] Ivan Vincent et.al. ,"Acute Lymphoid Leukemia Classification using Two-Step Neural Network Classifier", Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop, pp 1-4, 2015

[55] Vasuki Shankar et.al. ,"Automatic detection of Acute Lymphoblastic Leukemia using Image Processing", 2016 IEEE International Conference, pp 186-189, 2016

 DOI : 10.18231/2454-9150.2019.0204