

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Leukemia Diagnosis using Back Propagation Multi-layer Feed Forward Neural Networks

Shahila K¹, Kavitha.S.N²

M.Tech Scholar, Dept. of IT, RVCE, Bangalore, India¹

Asst. Professor, Dept. of IT, RVCE, Bangalore, India²

ABSTRACT: The term “Leukemia” refers to cancer of the blood cells. The presence of Eosinophil, Neutrophil, Lymphocyte, Basophil, and Monocyte in the blood test and the bone marrow indicates that the cells are influenced by leukemia. The diagnosis of ailments in blood plays a crucial role in the appropriate discovery and curative of diseases and thus protection and being saviour to the human life. Automatic detection of the ailment tends to reduce the erroneous results obtained from the microscopic images and also reduces the occurrence of mistakes by the pathologists. For automatic detection of leukemia the blood smear microscopic images are taken as an input. Microscopic images are pre-processed to remove the noise. Images are segmented for classification into appropriate structures. Otsu threshold method is used for Image segmentation. Dynamic features like area, ellipse, circularity, arch length, contour convex are used to identify the candidates of the images. Images are classified as containing the components Eosinophil, Neutrophil, Lymphocyte, Basophil, and Monocyte in the microscopic images using Back propagation for multi-layer Feed forward *neural networks is used for training the images. Accuracy of 85% is achieved in Leukemia classification.*

KEYWORDS: Otsu Threshold, Microscopic images, Feature Extraction, Segmentation, Back propagation, Multi-layer perceptron, Feed-Forward

I. INTRODUCTION

Leukemia is a malignant disease in which abnormal leucocytes is produced. The presence of Eosinophil, Neutrophil, Lymphocyte, Basophil, and Monocyte in the blood test and the bone marrow indicates that the cells are influenced by Leukemia. Processing of images and analysing the same contributes a way to excerpt and enumerate the objects and recognition of patterns in images.

Microscopic images are pre-processed to remove the noise. Images are segmented to classify the images into appropriate structures. Otsu threshold method is used for Image segmentation. Threshold method is used to generate binary images from the gray scale images. Thresholding is an iterative process, with thresholding method computation of the spread of the pixels around the threshold are calculated. Features are extracted to classify the input data into various sets of features. Dynamic features like area, ellipse, circularity, arch length, contour convex are used to identify the candidates of the images. Images are classified as containing the components Eosinophil, Neutrophil, Lymphocyte, Basophil, and Monocyte in the microscopic images using artificial neural networks. Microscopic images are multifaceted and strident. The diagnosis of ailments in blood plays a crucial role in the appropriate discovery and curative of diseases and thus protection and being saviour to the human life. Back Propagation is the method that is used for training. Back Propagation is used in multi-layer feed forward networks to train the leukemia images. Multilayer indicates multiple hidden layers. The nodes are called hidden neurons. Neurons extract higher order statistics by considering the external input which is given to the network and also the network output. Figure 1 depicts neural networks diagram of Leukemia diagnosis system.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

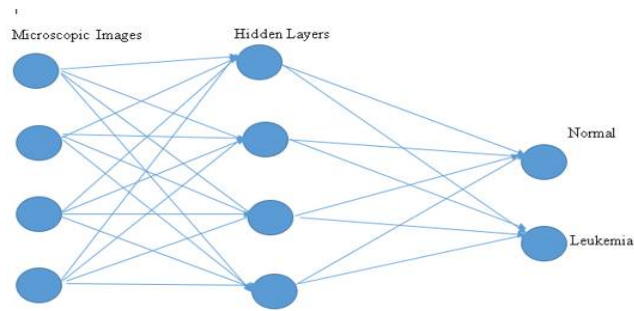


Figure 1-Leukemia diagnosis system

II. LITERATURE REVIEW

X. CHEN and colleagues in their research work have proposed a novel method for segmentation and classification of leukocyte nuclei. This method is used by the biologist's. In this method segmentation and classification of leukocyte nuclei is performed by which, the type of leukocytes texture and shape is detected. A cellular module known as white corpuscle and hemoglobin is absent in the white corpuscle. It has a nucleus and it acts as an immune to the body, it protects the body against infections. It acts as an immunity to the body by producing the antibodies [1]

J. Kawahara and colleagues in their research work has proposed diagnosing leukemia using classifiers and using CNN [Convolutional Neural Networks] on the pre-trained data. The approach proposed in this paper describes diagnosis of Leukemia without the segmentation process. In this work the evaluation is one by using pre trained CNN's such as Alex Net, CaffeNet and Vgg-f. Along with these CNN's, feature vectors are also considered. For extracting the features Vgg-f is considered. PCA [Principal Component Analysis] is used to choose out the features as the final description. Combination of 3 classifiers are proposed SVM[Support Vector Machine], MLP [Multilayer Perceptron] and (RF) [Random Forest]. The proposed system uses a blood smear image data base that contains multiple nuclei for every image.[2]

III. METHODOLOGY OF THE PROPOSED SYSTEM

The methodology followed in the proposed system is depicted in the figure 2.

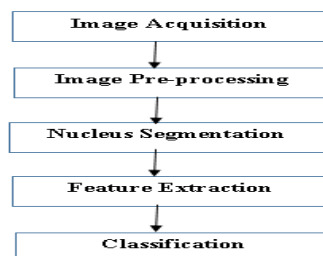


Figure 2. Methodology of Leukemia diagnosis system

Image Acquisition - The images from the blood smear and the bone marrow are acquired as the first step. The database for the study of Leukemia is prepared as follows. The data set is taken from MIT data base. The blood smear images of 367 leukaemia influenced patients are considered. 27 normal blood smear images are considered. High resolution digital camera from Nikon Eclipse under 100X was used to take the microscopy of this images. Images have the maximum resolution pixels and it is in .jpg format. Haematologist have determined the appropriateness of the

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

images and the blood cell type. For the purpose of this research 607 Leukemia blood cell images are acquired and 597 normal blood cell images are acquired.

Image Pre-processing- For pre-processing image enhancement is used as a first step. Filtering is used to remove the noise. The pre-processing step is used to improve the quality of the image and to figure out the region of interest that is the nucleus of the image. The robustness of the approach can be improved and fluctuations can be removed by applying a histogram equalization function. The image is converted from RGB to H, S and V values. Correlation between the channels can be reduced. The converted image can be calculated with separate H, S and V values. As they treated as separate channels they are not correlated. The HSV can be expanded as H as hue S as saturation and V as intensity. Saturation represents tint in color. This is human perception of lightness. In the software the color is considered as hue and saturation. Gray level of intensities of the image is equalized by applying Histogram equalization on V channel. In different image acquisition conditions the effects of color levels is lightened. Figure 3a,3b depicts the original RGB image and converted to HSV image.

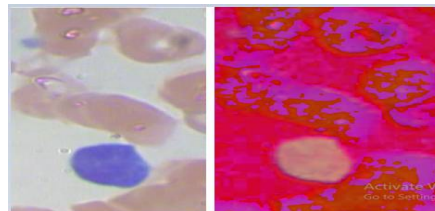


Figure 3a. RGB image 3b. HSV image

Feature Extraction- Area, perimeter centroid, circularity are the characteristics of lymphocyte cells that is used to differentiate the blasts. Area of the WBC is calculated first and then the area of the WBC nucleus is considered. The number of pixels in the WBC nucleus is calculated.

- i. Area – Area is calculated by considering the non-zero pixels and the pixels within the image.
- ii. Circularity-It is not having any dimension it keeps changing according to the surface irregularities
- iii. Eccentricity-This is a measurement to figure out the deviation of the normal cells from being circular. The value is either 0 or 1
- iv. Solidity-It is the density calculation. If the obtained value is less than 1 then the boundaries are irregular.

Classification- Back Propagation for multi-layer feed forward network is used for classification. *Multilayer indicates multiple hidden layers. The nodes are called hidden neurons. Neurons extract higher order statistics by considering the external input which is given to the network and also the network output.*

K hidden layers are considered in a multi-layer perceptron input layer + output layer =K+2 layers.

Input=K output=L hidden layer units = Nm

Weight of ith and jth unit of layer m= w_{ij}^m

m-layer'sith unit activation= x_m^i ,

x_m^0 is an input and for k+1 of m layers x_m^i is an output.

$$u(n)=(x_1^0(n), \dots, x_k^0(n))^t, d(n)=(d_1^{k+1}(n), \dots, d_L^{k+1}(n))^t \rightarrow 1$$

n represents the training instance. d is the training data. T is the input and output vector valued pairs. u (n) is the training input. Non-input units activation

$$x_i^{m+1}(n) = f\left(\sum_{j=1, \dots, N^m} w_{ij}^m x_j(n)\right)$$

Equation 1 is run to obtain the activation for subsequent hidden layers until a response from the network is obtained at the output layer. Network response is represented as y

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

$$y(n) = (x_1^{k+1}(n), \dots, x_L^{k+1}(n))$$

Summed square error is represented as E

$$E = \sum_{n=1, \dots, T} \|d(n) - y(n)\|^2 = \sum_{n=1, \dots, T} E(n)$$

The objective of training the data is to reduce the summed square error. Reduction of the errors is possible by changing the weights in the error gradient direction by considering the respective weights. Weights is represented by w.

$$\frac{\partial E}{\partial w_{ij}^m} = \sum_{t=1, \dots, T} \frac{\partial E(n)}{\partial w_{ij}^m}$$

Learning rate= γ

$$\text{New } w_{ij}^m = w_{ij}^m - \gamma \frac{\partial E}{\partial w_{ij}^m} \rightarrow 2$$

New weights are obtained by executing equation 2 in batch learning mode .One pass of training samples is called as epoch. Before the epoch begins weights are initialized to some random numbers. These weights are changed after individual training samples.

$$w_{ij}^m = w_{ij}^m - \gamma \frac{\partial E(n)}{\partial w_{ij}^m}$$

$$\text{Error gradient} = \frac{\partial E(n)}{\partial w_{ij}^m}$$

After every level of epoch the error is computed. This process is run in an iteration until the error falls below the threshold or it is exceeding from the number of epochs defined.

1 epoch=O (T M) where m is the total network connections.

T represents addition and multiplication.

Figure 4 depicts the architecture diagram of Leukemia diagnosis system

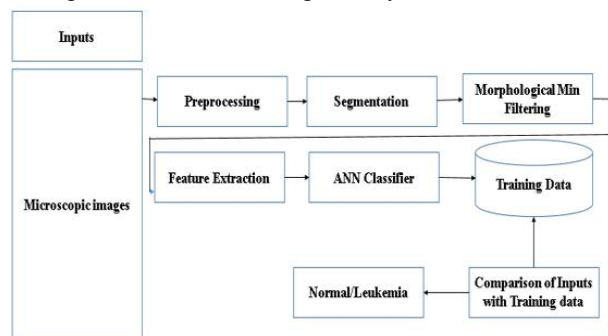


Figure 4. Architecture Diagram Leukemia diagnosis system

IV. IMPLEMENTATION

This section presents steps involved in various algorithms used in various steps of Leukemia diagnosis system.

Algorithm for Leukemia identification

Algorithm: Detecting the abnormalities in Retinal Fundus Images

Input: Microscopic Images to detect Leukemia

Output: Detection of Leukemia influenced cells.

Step 1: Microscopic images

Step 2: Pre-Processing of the microscopic image

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Image enhancement to increase clarity

Filtering to remove the noise

Step 3: Segmentation using Otsu's threshold m

Probability of the intensity of the pixel is considered and histogram is computed.

Initial values $\omega_i = 0$, $\mu_i = 0$, t

For all possible threshold values $t=1 \dots \dots$ maximum

Calculate $\sigma_b(t)$

1. The threshold obtained corresponds to the maximum value of $\sigma_b(t)$.

Step 5: Candidate Extraction

Dynamic shape features to discriminate between false and true characteristic

- i. Area – Area is calculated by considering the non-zero pixels and the pixels within the image.
- ii. Circularity-It is not having any dimension it keeps changing according to the surface irregularities
- iii. Eccentricity-This is a measurement to figure out the deviation of the normal cells from being circular. The value is either 0 or 1
- iv. Solidity-It is the density calculation. If the obtained value is less than 1 then the boundaries are irregular

Step 6: Classification using Back propagation multilayer feed forward classifier

Extract the bigger cells

Extract the smaller cells

Classify the cells as per the images

If image = 'Eosinophil'

$E=e+1$

If image = 'Lymphocyte'

$L=l+1$

If image = 'Basophil'

$B=b+1$

If image = 'Neutrophil'

$N=n+1$

End

V. RESULTS

To implement the Proposed System Python is used Automatic detection of different types of leukaemia such as Eosinophil, Lymphocyte, Basophil, Neutrophil and Monocyte will reduce the errors that occurs due to the erroneous reading of microscopic images. The dataset is taken from MIT database. 1204 microscopic images are learnt using neural networks. 367 images are considered as test data. 311 images are classified correctly that is the diagnosis of classification of leukaemia is correct for 311 images. Accuracy of 85% is achieved in classifying the images. Back propagation multilayer feed forward classifier provides greater accuracy and lower risk of human errors. Figure 5 to Figure 13 depicts the various classification of leukaemia by examining the presence of the different components like Eosinophil, Lymphocyte, Basophil, Neutrophil and Monocyte in the microscopic images.

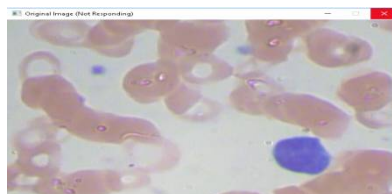


Figure 5. Lymphocyte component in the microscopic image

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Figure 5 shows the Leukemia component Lymphocyte which is present in the blood smear test. The presence of the lymphocyte component diagnoses the blood smear test as leukaemia.

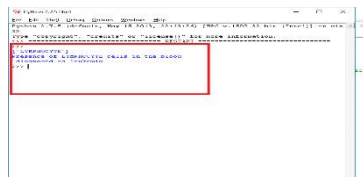


Figure 6. Presence of lymphocyte, diagnosed as Leukemia

Figure 6 shows the confirmation message on the screen that lymphocyte component is recognised and it is diagnosed as Leukemia

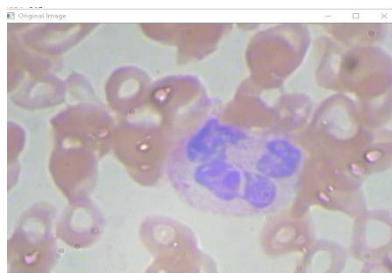


Figure 7. Neutrophil component in the microscopic image

Figure 7 shows the Leukemia component Neutrophil which is present in the blood smear test. The presence of the Neutrophil component diagnoses the blood smear test as leukaemia.



Figure 8. Presence of neutrophil, diagnosed as Leukemia

Figure 8 shows the confirmation message on the screen that neutrophil component is recognised and it is diagnosed as Leukemia

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018



Figure 9. Basophil component in the microscopic image

Figure 9 shows the Leukemia component Basophil which is present in the blood smear test. The presence of the Basophil component diagnoses the blood smear test as leukaemia.



Figure 10. Presence of Basophil, diagnosed as Leukemia

Figure 10 shows the confirmation message on the screen that Basophil component is recognised and it is diagnosed as Leukemia

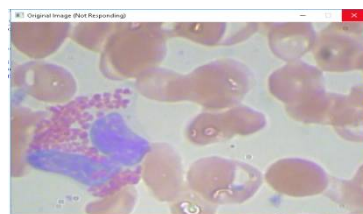


Figure 11. Eosinophil component in the microscopic image

Figure 11 shows the Leukemia component Eosinophil which is present in the blood smear test. The presence of the Eosinophil component diagnoses the blood smear test as leukaemia.



Figure 12. Presence of Eosinophil, diagnosed as Leukemia

Figure 12 shows the confirmation message on the screen that Eosinophil component is recognised and it is diagnosed as Leukemia

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

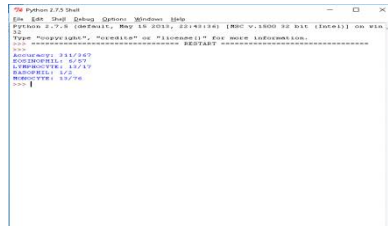


Figure 13. Accuracy of classification of images. [Classifier Analysis]

Figure 13 shows the results of classifier analysis. Accuracy of 85% is achieved.

VI. CONCLUSION

This Paper presents the design and development of Leukemia diagnosis system. Leukemia is a malignant disease in which abnormal leucocytes are produced. The presence of Eosinophil, Neutrophil, Lymphocyte, Basophil, and Monocyte in the blood test and the bone marrow indicates that the cells are influenced by leukaemia. Microscopy image is pre-processed to remove the noise. Images are segmented to classify the images into appropriate structures. Otsu threshold method is used for Image segmentation. Dynamic features like area, ellipse, circularity, arch length, contour convex are used to identify the candidates of the images. Images are classified as containing the components of Eosinophil, Neutrophil, Lymphocyte, Basophil, and Monocyte in the microscopic images using neural networks. Accuracy of 85% is achieved in classification of the images. Back propagation multilayer feed forward classifier provides more accuracy and less errors.

REFERENCES

1. X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), vol. 37, August 2015, pp. 715–718.
2. J. Kawahara and G. Hamarneh, Multi-resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers. Springer International Publishing, 2016.
3. D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," ArXiv e-prints, June 2016.
4. M. Madhukar, S. Agaian, and A. T. Chronopoulos, "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," IEEE Systems Journal, vol. 8, no. 3, pp. 995–1004, 2014.
5. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
6. S. Mohapatra, D. Patra, and S. Satpathi, "Image analysis of blood microscopic images for acute leukemia detection," in Proc. IECR, 2010, pp. 215–219.
7. Ravichandran K. S. and Ananthi B. (2009), Color Skin Segmentation using K-Means Cluster. International Journal of Computational and Applied Mathematics, Vol. 4, No.2, pp.153 – 157
8. Marina E. P., Christophoros Nikou., and Antonia Charchanti. (2011), "Automated Detection of Cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering," IEEE Transactions on Information Technology in Biomedicine, Vol. 15, No. 2.
9. Barbara JB. A Beginner's Guide to Blood Cells. India: Blackwell Publishing Ltd.; 2004. p. 64–5.
10. Mohapatra S, Patra D. Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images. In Systems in Medicine and Biology (ICSMB), 2010 International Conference on, IEEE; 2010.
11. Mokhtar N, Harun NH, Mashor MY, Rosline H, Adollah R, Adilah N, et al. Image Enhancement Techniques Using Local, Global, Bright, Dark and Partial Contrast Stretching For Acute Leukemia Images. Proceedings of the World Congress on Engineering 2009 Vol I WCE 2009, July 1–3, 2009, London, U.K.
12. Rodenacker K, Bengtsson E. A feature set for cytometry on digitized microscopic images. Anal Cell Pathol 2003;25:1–36