

Blood Leukemia Classification using Discrete Wavelet Transform and Artificial Neural Network

Bharti Choudhary
M.Tech Scholar (ECE)
CIIT, Indore
choudhary_vikas007@yahoo.co.in

Santosh Pawar
Associate Professor (ECE)
CIIT, Indore
spawarrkdf@gmail.com

Sheetesh Sad
Asst.Professorr (ECE)
CIIT, Indore
sheeteshsad@gmail.com

Abstract: This paper presents an efficient mechanism for the classification of blood leukaemia using microscopic images employing a Probabilistic Neural Network (PNN). Probabilistic Neural Network is based on Bayes's theorem of Conditional Probability and is a famed paradigm for data classification for systems employing artificial intelligence. Pre-processing has been achieved using gray scaling and thresholding. Discrete Wavelet Transform (DWT) has been used as a tool to remove the abrupt variations in the calculated feature values. Principal component analysis (PCA) has been used to find particular trends in the computed feature data and minimize the redundancy. It has been shown that the proposed technique achieves 97% percent classification accuracy which can be attributed to the highly rigorous pre-processing and feature extraction mechanism which culminates in training a Probabilistic Neural Network. The category of ANN used here for classification is the Probabilistic Neural Network which is highly effective in classification problems.

Keywords: *Artificial Intelligence (AI), Probabilistic Neural Network (PNN), Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), Accuracy, Leukaemia.*

I. INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network (ANN) has recently emerged as one of the most powerful tools for contemporary computation. Its design is based on the fact that the human brain is a:

- 1) **Highly Non Linear Structure**
- 2) **Highly Parallel Structure**

3) **Adaptive Learning without the need for explicit programming.**

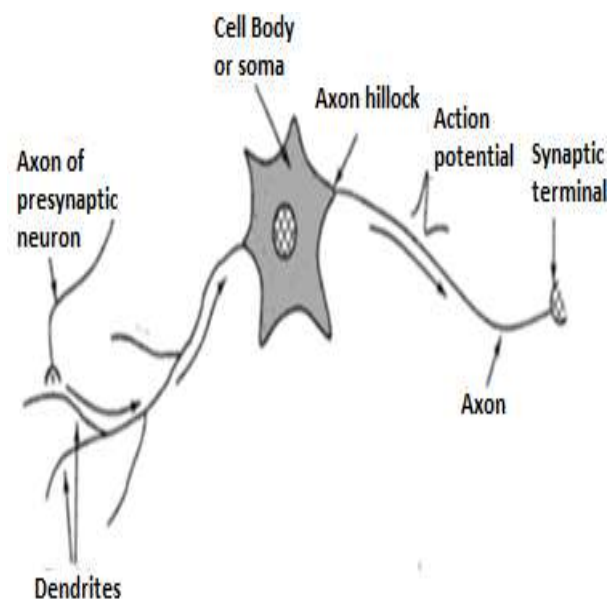


Fig.1 Biological Model of a Single Neuron

The analysis of the biological model of the neuron helps us in designing the mathematical model of the neural network. An extremely important attribute of the neural model is its ability in following trends in the input fed to it. No matter how complex or abruptly the output for a corresponding input may change, the network maps the input and output in the form of experiences called weights. The network continuously learns and adapts according to the provided input data and its corresponding data. Thus it can be thought of working as the human brain that also adapts itself according to experiences. The experiences themselves have counterparts in the mathematical ANN structure called weights. The parallel structure enables data or inputs X from various paths design the weights W . The design of

the network culminates in the decision making according to some function θ called the bias. The structure can be mathematically modelled as:

$$Y = \sum_{i=1}^n X_i W_i + \Theta \quad (1)$$

Here X represents the signal

W represents the weight

Θ represents the bias.

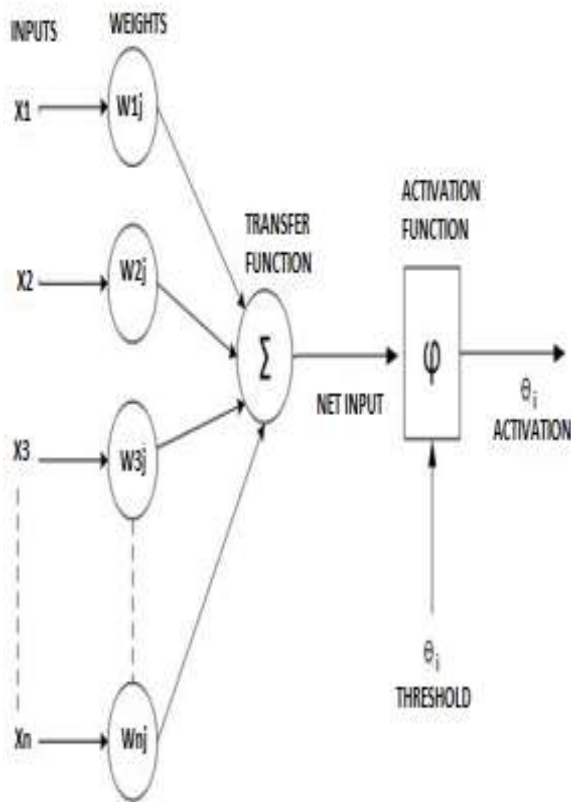


Fig.2 Mathematical Formulation of Neural Network

The above figure is pretty generic in nature. The task of following a trend in the given input data can be accomplished accurately using a particular architecture of a neural network. If the correspondance between the input X and a target Y is given to the network to design and adapt its weights, then the deviation of the expected output and the actual output is given by an error E which can be mathematically defined as:

$$E = Y_a - Y_p \quad (2)$$

Y_a is the actual output

Y_p is the predicted output

E is the error.

II. PROBABILISTIC NEURAL NETWORK (PNN)

PNN is built on the theory of Bayesian network and the estimation of probability density function. This theory allows for cost function to represent the fact that it may be worse to misclassify a vector that is actually a member of class A but it's classified as a vector that belongs to class B. PNN is extensively used for classification problem. It works as a supervised classifier. We give a sample of our data set to it as an input and the data is then handled through numerous layers. The Baye's rule such that the input vector belonging to class A is classified as

$$P_A C_A f_A(x) > P_B C_B f_B(x) \quad (3)$$

Where P_A –Priori probability of occurrence of pattern in class A. C_A - Cost associated with classifying vectors.

$f_A(x)$ is the Probability density functions of class A. PNN can take any number of input and can map them to any number of output so for that reason PNN works more efficiently than other Neural Networks. Neural Networks in common have a back propagation which this neural network doesn't have making it faster than others.

The different layers of Probabilistic Neural Network are:

- **Input Layer:** It consists of the input to the neural network the training sample whose class has to be found out.
- **Pattern Layer:** It consists of the Gaussian function with the training samples are the center to it. We then find the Gaussian distance for all of them using the formula.
- **Summation Layer:** Here all the output of the pattern layer is received and then they are added and send to the output layer for class determination. It is also the hidden layer in the Neural Network.
- **Output Layer:** Here all the outputs from the summation layers are taken and a selection of the

largest value is done and that particular class of the selection determines the class of the sample.

The PNN works by creating a set of multivariate probability densities that are derived from the training vectors presented to the network. The input instance with unknown category is propagated to the pattern layer. Once each node in the pattern layer receives the input, the output of the node will be computed. Here d is the number of features of the input instance x , σ is the smoothing parameter, and x_{ij} is a training instance corresponding to category c . The summation layer neurons compute the maximum likelihood of pattern x being classified into c by summarizing and averaging the output of all neurons that belong to the same class

$$\Pi_i^c = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}\right] \quad (4)$$

where N_i denotes the total number of samples in class c . If the a priori probabilities for each class are the same, and the losses associated with making an incorrect decision for each class are the same, the decision layer unit classifies the pattern x in accordance with the Bayes's decision rule based on the output of all the summation layer neurons

$$C(x) = \operatorname{argmax} \{P_i(X)\}, i = 1, 2, \dots, c \quad (5)$$

where $C(x)$ denotes the estimated class of the pattern x and m is the total number of classes in the training samples. If the a priori probabilities for each class are not the same, and the losses associated with making an incorrect decision for each class are different, the output of all the summation layer neurons will be

$$C(x) = \operatorname{argmax} \{P_i(X) \operatorname{cost}_i(x) \operatorname{apro}_i(x)\}, i = 1, 2, \dots, c \quad (6)$$

Where $\operatorname{cost}_i(x)$ is, the cost associated with misclassifying the input vector and $\operatorname{apro}_i(x)$. Thus PNN acts as an effective classifier to data classification into categories.

There are quite a lot of advantages of using PNN instead of Back Propagation (BP) or multilayer perceptron. PNNs are much faster than BP multilayer networks. It provides better accuracy.

This network is relatively insensitive to outliers and generates accurate predicted target probability scores.

Some of the advantages of PNN are:

- It can map any number of input to any number of classification.
- Unlike other Neural Network it doesn't have a back propagation which makes it a fast learner.
- Once trained removing or adding a data sample doesn't require retraining of the sample.
- It saves a lot of computational space.

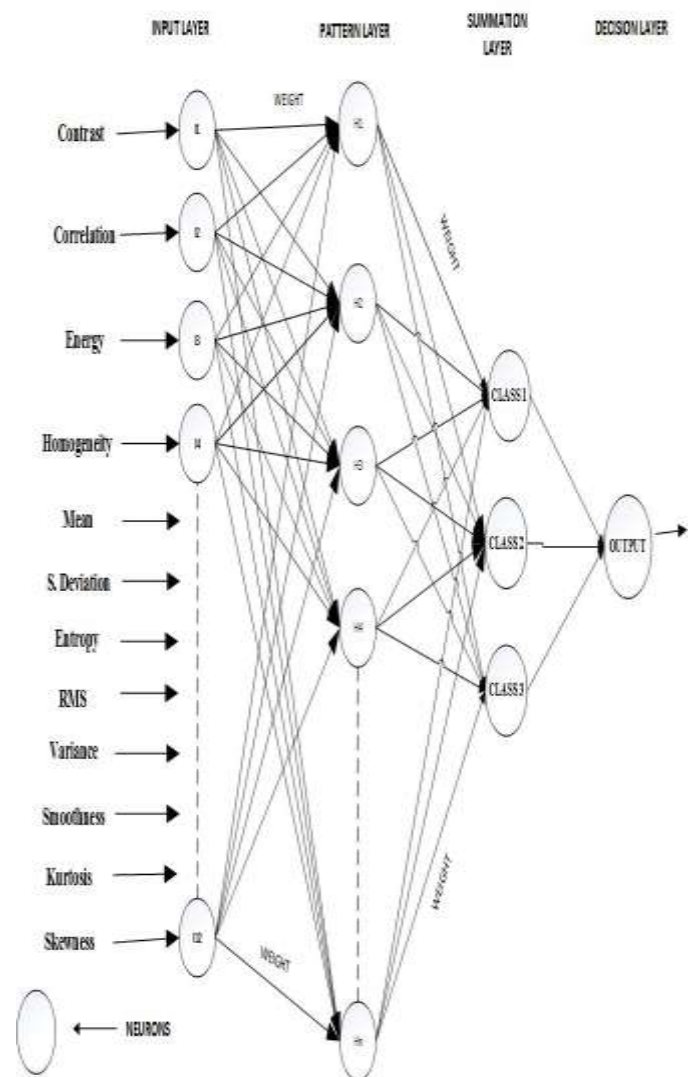


Fig.3 Model of PNN with Input Feature Values

III. DATA PRE-PROCESSING

Prior to feature extraction and classification, data pre-processing is done to achieve train the PNN

accurately and obtain high sensitivity and accuracy of classification.

(i) Segmentation:

The division of an image into meaningful structures, image segmentation, is often an essential step in image analysis, object representation, visualization, and many other image processing tasks. A disjunct categorization does not seem to be possible though, because even two very different segmentation approaches may share properties that defy singular categorization. The categorization presented in this chapter is therefore rather a categorization regarding the emphasis of an approach than a strict division. The following categories are used:

Threshold based segmentation: Histogram thresholding and slicing techniques are used to segment the image. They may be applied directly to an image, but can also be combined with pre- and post-processing techniques.

Edge based segmentation: With this technique, detected edges in an image are assumed to represent object boundaries, and used to identify these objects.

Region based segmentation: Where an edge based technique may attempt to find the object boundaries and then locate the object itself by filling them in, a region based technique takes the opposite approach, by (e.g.) starting in the middle of an object and then “growing” outward until it meets the object boundaries

Segmentation plays a crucial role in the feature extraction and classification.

(ii) Principal Component Analysis (PCA)

Principal Component Analysis is a method by which we can find the attributes which contribute the most to the final classification of the data set. Principal component analysis (PCA) is a mathematical procedure that converts a set of correlated variables into a set of values of uncorrelated variables called principal components.

PCA is a mathematical process that does a linear transformation of the dataset to a new system such that the element which influences to the class greatest is set as the first coordinate, that is the first principal component followed by the next great and so on. But when the data is non-linear in nature, PCA is unable to find the principal components.

(iii) Discrete Wavelet Transform (DWT)

The Wavelet Transform is rather a recent tool for the analysis of randomly fluctuating non-smooth signals. The mathematical description of the wavelet transform can be given by:

$$C(S, P) = \int_{-\infty}^{\infty} f(t) ((S, P, t)) dt \quad (7)$$

Here S stands for scaling

P stands for position

t stands for time shifts.

C is the Continuous Wavelet Transform (CWT)

The main disadvantage of the CWT is the fact that it contains an enormous amount of data. The sampled version of the CWT is the Discrete Wavelet Transform (DWT). The DWT is a down sampled version of the CWT and its characteristic nature is to smoothen out abrupt fluctuations which are possible due to both abruptly changing base functions and down sampling.

The scaling function can be defined as:

$$W\Phi(J_0, k) = \frac{1}{\sqrt{M}} \sum_n S(n). \Phi(n)_{j_0'k} \quad (8)$$

The Wavelet function can be defined as:

$$W\Psi(j, k) = \frac{1}{\sqrt{M}} \sum_n S(n). \psi(n)_{j,k} \quad (9)$$

Where $\frac{1}{\sqrt{M}}$ is Normalizing term

IV. FEATURE EXTRACTION

Since the neural network does not understand images, rather it understands numerical data. At the outset the images pre-processed thus far should be converted to assertive numerical values. Subsequently, important parameters or features should be calculated to train and finally test the neural network for classification. The vital parameters of the leukaemia images are also called features. Prior to feature extraction though, it is necessary to pre-process the image data to attain correct values of the features

The various features calculated for training the neural network are:

- Contrast
- Correlation
- Energy
- Homogeneity
- Mean
- Standard Deviation
- Entropy
- RMS
- Variance
- Smoothness
- Kurtosis
- Skewness

V. ALGORITHM USED

The proposed technique can be understood using the following steps of the algorithm:

- 1) One image is loaded at a time to the MATLAB workspace.
- 2) RGB to Gray scale Conversion and Binary Thresholding to convert the image into a Binary image have been employed as a pre-processing tool.

- 3) Segmentation has been applied to locate affected cellular area.
- 4) Subsequently Discrete Wavelet Transform (DWT) has been applied smoothen out the non-linearities in the image to facilitate feature extraction.
- 5) Principal Component Analysis (PCA) has been employed to find regular patterns or trends in the DWT coefficients computed above.
- 6) Twelve feature values have been calculated for the image under consideration.
- 7) The above process is employed for all the training data samples in the data set.
- 8) The feature values obtained for all the training data samples are used to train a Probabilistic Neural Network (PNN).
- 9) The Neural Network is tested for the remaining images of the data set.
- 10) Sensitivity and Accuracy have been evaluated for the Proposed System Design.

The working of the proposed algorithm can be understood as follows:

The leukemia images are microscopic images and show extremely minute variation to be recognized by the naked eye. Hence computer aided classification is needed. An effective technique is training a Probabilistic Neural Network using the training data set of Leukemia and Normal Blood images. The important point here to note is the fact that neural networks do not understand or work upon images or data types other than numbers. Hence the most important features of images are computed (which are numerical values) and a PNN is trained. After training is done, the PNN adapts its weights accordingly. **Classification mechanism of PNN:** After the PNN is trained by an exhaustive set of pre-processed feature values, new images not used for training are used for testing. Prior to testing, the features of any new image in the testing data set are computed to be fed to the PNN. The following flowchart illustrates the process.

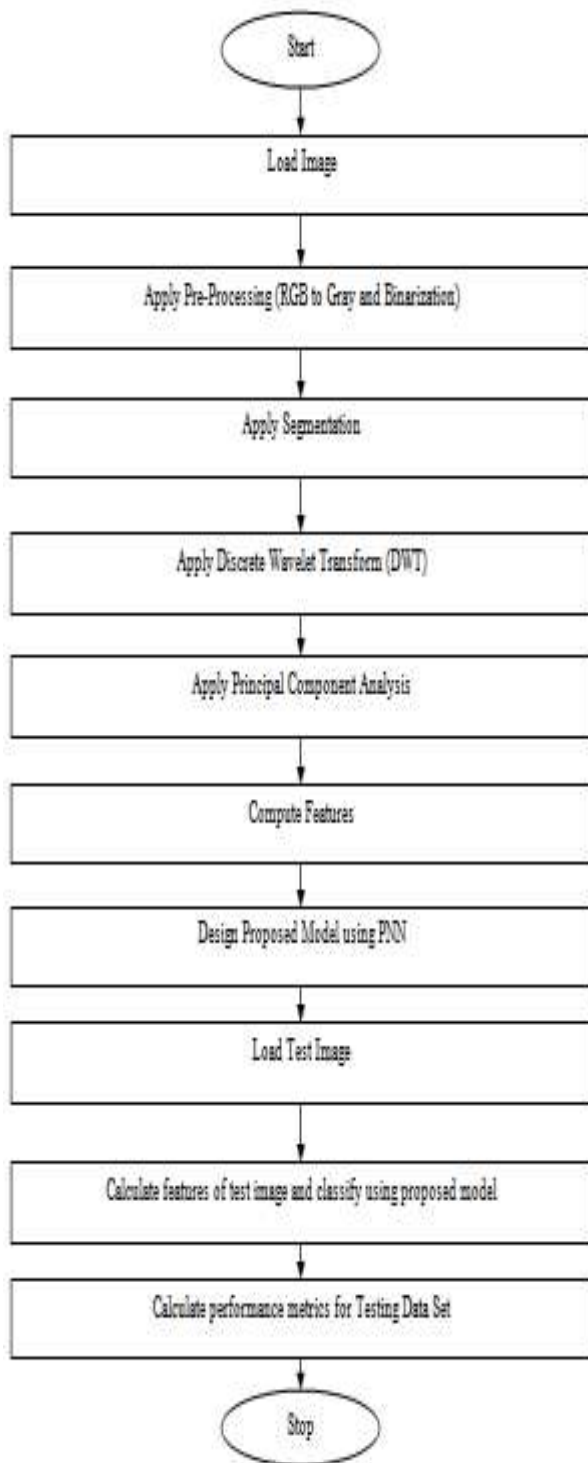


Fig.4 Flowchart of Proposed Algorithm

It is worth mentioning that pre-processing of the training data is critical since it decides how the Neural Network will classify later on. The classification stage used RGB to Gray conversion, Binarization and Discrete Wavelet Transform (DWT). Further Principal Component Analysis (PCA) is used to remove redundancies in feature

values and train the neural network with only those feature values which affect the outcome the most. As a standard convention, almost 70% of the data is used for training and remaining 30% is used for testing. A total of 260 images have been used out of which 200 images have been used for training and 60 are used for testing.

VI. RESULTS:

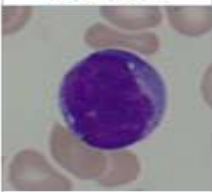

Evaluation Parameters:

The various parameters for the classification are:

1. **True Positive (TP):** It is the case when a sample belongs to category and the test also predicts its belongingness.
2. **True Negative (TN):** It is the case when a sample does not belong to category and the test also predicts its non-belongingness.
3. **False Positive (FP):** It is the case when a sample does not belong to category and the test predicts its belongingness.
4. **False Negative (FN):** It is the case when a sample belongs to category and the test predicts its non-belongingness.

Accuracy (Ac): It is mathematically defined as:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

HEALTHY	LEUKEMIA
	
0.566477273	0.403409091
0.134990779	0.079914238
0.805264075	0.689903151
0.938386769	0.910672348
0.004660435	0.009083011
0.106527261	0.106241438
2.208658726	3.764806736
0.106600358	0.106600358
0.011328815	0.011261729
0.895968612	0.943774119
20.05759459	7.15389322
2.159108057	0.788548077

$$= \frac{29+29}{29+29+1+1} \cdot 100 = 97\%$$

The results show significant improvement compared to the previous work, (Computer Aided Diagnostic System for Detection of Leukemia using Microscopic Images, by Jyoti Rawat et.al, Elsevier, Science Direct 2015), which achieves an accuracy of 89.8%

Conclusion: In this paper blood leukaemia image are automatically classified into normal or malignant (with leukaemia). The classifier used is the Probabilistic Neural Network (PNN) classifier. Accuracy of 97% is obtained. This work will act as supportive tool for medical practitioners and will help doctor for fast diagnosis based on which the treatment plan can be decided. It has been found that the proposed technique achieves 97% classification accuracy which can be attributed to the efficacy of the proposed method.

Fig.5 :Tabulation of featue values for Normal and Leukaemia Blood samples

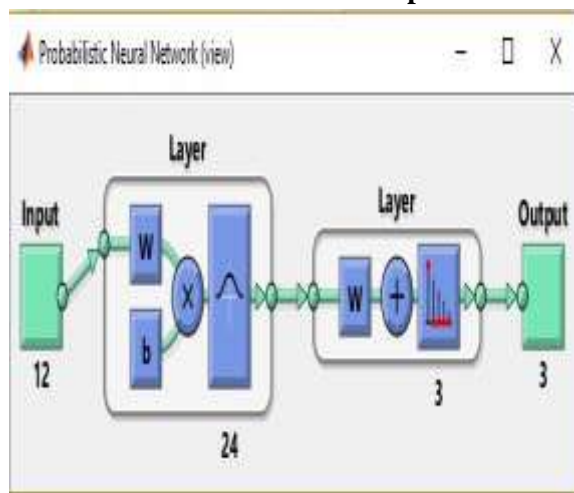


Fig.5 :Designed PNN

The designed Probabilistic Neural Network is shown above, which has 12 inputs corresponding to the 12 feature values.

A total of 200 images have been used for training and 60 images have been used for testing. It has been found that the accuracy is 97%.

The calculation has been done as follows:

Out of the testing dataset of 60 images,

FP=1, FN=1.

Therefore,

$$\text{Accuracy } Ac = \frac{TP+TN}{TP+TN+FP+FN}$$

References

- [1] Classification of Leukemia Blood Samples Using Neural Networks, by Malek Adjouadi, Melvin Ayala, Mercedes Cabrerizo, Nuannuan Zong, Gabriel Lizarraga, Mark Rossman Springer 2010
- [2] Classification of blasts in acute leukemia blood samples using K-Nearest Neighbor, by N.Z Supardi, M.Y Mashor, N.H Harun, F.A. Bakri, R.Hassan, IEEE 2011
- [3]Detection of Leukemia in Microscopic Images Using Image Processing, by Chaitali Raje, Jyoti Rangole, IEEE 2014
- [4] Computer Aided Diagnostic System for Detection of Leukemia using Microscopic Images, by Jyoti Rawat, Annapurna Singh, H S Bhadauria, Jitendra Virmani, Elsevier 2015
- [5] Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine by S. Ravikumar, Elsevier 2015
- [6] Large Scale Image Feature Extraction from Medical Image Analysis, by Desai Devanshi Manojbhai and R.Rajamenakshi IJAERS 2016
- [7] Efficacy of Azacitidine in autoimmune and inflammatory disorders associated with myelodysplastic syndromes and chronic myelomonocytic leukemia by Jean Baptiste et.al, Elsevier 2016
- [8] Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images: A Comprehensive Review, Fuyong Xing and Lin Yang, IEEE 2016by
- [9] Detection of Blood Cancer in Microscopic Images of Human Blood Samples: A Review, M.Saritha,Prakash.B.B,Sukesh.K and Shrinivas. B, IEEE 2016
- [10] Design of New Architecture to Detect Leukemia Cancer from Medical Images,by Gurpreet Singh,Gaurav Bathla and SharanPreet Kaur IJAER 2016
- [11] Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia, Esti Suryani, Wiharto, and Nizomjon Polvonov, IJCNS 2015

- [12] Leukemia Detection using Digital Image Processing Techniques, by Himali P. Vaghela, Hardik Modi, Manoj Pandya, M.B. Potdar IJAIS 2015
- [13] Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features, Rajesh Kumar, Rajeev Srivastava, and Subodh Srivastava, Hindawi 2015
- [14] Segmentation of Large Scale Medical Images using HPC: Classification of Methods and Challenges by Kodrani Kajal Pradipkumar, R. Rajamenakshi, IJAERS 2015
- [15] Application of stationary wavelet entropy in pathological brain detection BY Shuihua Wang, Springer 2016
- [16] Biological implications of somatic DDX41 p.R525H mutation in acute myeloid leukemia by Moe Kadono, Elsevier 2016
- [17] Leucocyte classification for leukaemia detection using image processing techniques by Lorenzo Putzua, Giovanni Caocobb, Cecilia Di Rubertoa, Elsevier 2014
- [18] Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images by Sos Agaian, , Monica Madhukar, Anthony T. Chronopoulos, IEEE 2014
- [19] Counting of wbc and rbc from blood images using gray thresholding pooja r. patil, g.s. sable, gauri anand gaonkar, IJRET 2014
- [20] Comparative analysis of segmentation algorithms for Leukocyte Extraction in acute Lymphoblastic Leukemia Images by Jyoti Rwat et.al IEEE 2014
- [21] Selection the best features for leukocytes classification in blood smear microscopic imagesomid sarrafzadeh, Research gate 2014
- [22] <https://en.wikipedia.org/wiki/Leukemia>
- [23] C.R., Valencio, M.N., Tronco, A.C.B., Domingos, C.R.B., "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.
- [24] R., Adollah, M.Y., Mashor, N.F.M, Nasir, H., Rosline, H., Mahsin, H., Adilah, "Blood Cell Image Segmentation: A Review", *Biomed2008, Proceedings 21*, 2008, pp. 141-144.
- [25] N., Ritter, J., Cooper, "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", *30th Australasian Computer Science Conference, Conference in Research and Practice in Information Technology*, Vol. 62, 2007, pp. 161-169.
- [26] 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.
- [27] Abdul Nasir, Mustafa N, Mohd Nasir, " Application of Thresholding in Determining Ratio of Blood Cells for Leukemia Detection" in the Proceedings of International Conference on Man-Machine system (ICoMMS) oct 2009 ,Batu Ferringhi, Penang, Malaysia.
- [28] Bhagyashri G Patil, Prof. Sanjeev N.Jain , "Cancer Cells Detection Using Digital Image Processing Methods" in International Journal of Latest Trends in Engineering and Technology".
- [29] ADNAN KHASHMAN, ESAM AL-ZGOUL, "Image Segmentation of Blood Cells in Leukemia Patients" in RECENT ADVANCES in COMPUTER ENGINEERING and APPLICATIONS ISBN: 978-960- 474-151-9
- [30] Fauziah Kasmin, Anton Satria Prabuwo, Azizi Abdullah "detection of leukemia in human blood sample based on microscopic images: a study", journal of theoretical and applied information technology 31st december 2012. vol. 46 no.2.
- [31] Farnoosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, Badrul Hisham Abdul Kahar, And M-Iqbal Saripan, "A Framework For White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing" in Shulin Li (ed.), *Biological Procedures Online*, Volume 11, Number 1 to the author(s) 2009
- [32] 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
- [33] Tara.Saikumar, B.K.Anoop, P.S.Murthy in "Robust Adaptive Threshold Algorithm based on Kernel Fuzzy Clustering on Image segmentation" in Natarajan Meghanathan, et al. (Eds): ITCS, SIP, JSE-2012, CS & IT 04, pp. 99–103, 2012. © CS & IT-CSCP 2012
- [34] Michael Barnathan, "Mammographic Segmentation Using Wave Cluster", open access algorithms ISSN 1999-4893, Algorithms 2012.
- [35] Xu Gongwen, Zhang Zhijun, Yuan Weihua, XuLi Na, "On Medical Image Segmentation Based on Wavelet Transform" in IEEE , ISBN 978-1-4799-4262-6, June 2014
- [36] Chinki Chandhok "A Novel Approach to Image Segmentation using Artificial Neural Networks and K-Means Clustering" in Chinki Chandhok / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 ,Vol. 2, Issue 3, May-Jun 2012, pp. 274-279
- [37] Ng, HP ong,S.H; Foong, K.W.C in "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm" IEEE , ISBN 1-4244-0069-4,2006
- [38] N., Ritter, J., Cooper, "Segmentation and Border Identification of Cells in Images of Peripheral Blood Smear Slides", *30th Australasian Computer Science Conference, Conference in Research and Practice in Information Technology*, Vol. 62, 2007, pp. 161-169
- [39] I., Cseke, "A Fast Segmentation Scheme for white Blood Cell Images", *IEEE*, 1992, pp. 530-533.
- [40] A., Khashman, E., Esam Al-Zgoul, "Image Segmentation of Blood Cells in Leukemia Patients", *Recent Advances in Computer Engineering and Applications*, ISBN: 978- 60-474-151-9, 2009, pp. 104-109.
- [41] N.I., Birndorf, J.O., Pentecost, J.R., Coakley, K.A., Spackman, "An Expert System to Diagnose Anemia and Report Results Directly on Hematology Forms", *Computers and Biomedical Research*, Vol. 29, 1996, pp. 16-26.
- [42] S., Mohapatra, D., Patra, S., Satpathi, "Image Analysis of Blood Microscopic Images for Leukemia Detection", *International Conference on Industrial Electronics, Control and Robotics, IEEE*, 2010, pp. 215- 219.
- [42] T. Markiewicz, S. Osowski, B., Marianska, L., Moszczynski, "Automatic Recognition of the Blood Cells of Myelogenous Leukemia Using SVM", *Proceedings of International Joint Conference on Neural Networks, Montreal* , Canada, July 31 – August 4, 2005, pp. 2496-2501
- [43] 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.
- [44] Athira Krishnan, Sree Kumar K, " A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images" , (IJCISIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7877
- [45] Mohapatra S., Patra D., Satpathi S., "Image Analysis of Blood Microscopic Images for Leukemia Detection", *International Conference on Industrial Electronics, Control and Robotics, IEEE*, pp. 215-219, 2010.
- [46] Virmani J., Kumar V., Kalra N., Khandelwal N., "Characterization of primary and secondary malignant liver