FAB Classification based Leukemia Identification and prediction using Machine Learning

Krishna Kumar Jha
Assistant Professor, Dept. of CSE
Calcutta Institute of Technology
Uluberia, Howrah, India
kkjha.visual@gmail.com

Prasanta Das
MCA(Final Yr), Dept. of MCA
Calcutta Institute of Technology
Uluberia, Howrah, india
pdas79409@gmail.com

Himadri Sekhar Dutta
Assistant Professor, Dept. of ECE
Kalyani Govt. Engineering College
Kalyani, India
himadri.dutta@gmail.com

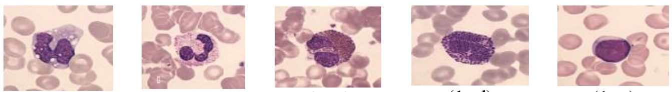
Abstract -- Background and Objective: Leukemia identification, detection, & classification has erupted an intriguing field in medical research. Several methodologies are convenient in theprevious work to detect five types WBCs (lymphocytes, eosinophils, monocytes, neutrophils, and basophils). Single cell Blood's smear images used for experiment. Propounded method is used for leukemia recognition, uncovering and distribution based on FAB classification.
Methodology: This propounded task has developed French-American and British (FAB) classification-based detection module on blood smearimages (BSIs). Methods like pretreatment, segmentation, feature extraction, distribution are used in detection method. The Propounded algorithm-based propounded model is used for segmentation, which is combination of the segmented results of the Linde-Buzo-Gray (LBG) algorithm, Adaptive canny used for edge identification and Hysteresis and watershed algorithm used for thresholding. The shape, texture features, color of segmented image are picked by neural network and classification is performed by Support Vector Machine (SVM) and prediction by Naïve Bayes Classifier (NBC).
Result: Dataset-master and Cellavison dataset is being used for the experimentation. The BSIs are considered for the Evaluation based on ROC curve analysis metrics like TPR, TNR and accuracy. Our propounded solution obtains superior classification performance in the given dataset. The suggested classifier enhanced the classification average accuracy to 99.06% and Mean Square Error (MSE) is 0.0407.
Conclusion: The enhanced accuracy had achieved by enhancing performance and classification with comparison with some other methods.

Keywords: WBCs; LBG; SVM; Naïve Bayes; neural network;

I. INTRODUCTION

Leukemia is treated as deadliest disease found in third world countries. Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML) are two types of Leukemia [1]. Identification of ALL and AML is tough task. Diagnosis of RBCs, platelets are more important rather than the organs. Evaluation of human blood is observed through microscopic BSIs [2]. Human interaction makes error prone to separate different WBC's nucleus as it is manual method. In rural areas uses of advance equipment is not feasible by the experts. To overcome these problems, automated segmentation techniques need to be developed. The automated segmentation of images has done through varying sizes and shapes [4-5]. Edge detection on cell images not performed perfectly because not all having sharp cell boundaries. It makes difficult to identify and locate all edge particular accurately [6]. Automated WBCs classification is performed by feature extraction. Features' characteristics plays important role for determining the accuracy rate by means of subsequent classification. WBCs' features can be classified into three categories: (a) texture feature, (b) Morphological feature and (c) color feature. Different features like circularity, size, shape of WBCs and nucleus can be included in morphological feature[7]. Individual color, texture feature of WBCs and chromatin and granule of nucleus serves unique and important feature for identification [8]. Standard deviation, kurtosis, mean, skewness, entropy of image can be included in texture feature [1]. Varience, mean, correlation analysis of the data of different classes are important features. Analysis of independent component used in application of linear SVM [9-10]. But features can be selected using neural network with SVM and NBC which demonstrated in this paper. The conventional method of feature extraction was outperformed by neural network.

In this paper our main propose is to developed French-American and British (FAB) classification-based detection method from BSIs. The segmentation will be done by propounded algorithm-based hybrid model, which produces segmentation by combining the results of Linde-Buzo-Gray algorithm, Adaptive canny, used for edge detection. Combination of Hysteresis and watershed algorithm is used for thresholding. Then features like texture, shape, color are extracted by proposed neural network using SVM and NBC for the classification.



(1_a) (1_b) (1_c) (1_d) (1_e)

Fig 1: five type of WBCs

II. LITERATURE SURVEY

Yampri P et al. [8] stated major reason for failure in edge detection because razor-edged cell boundaries make difficult to locate the cell accurately. Feature extraction process for automated WBCs classification relies on morphological features, color features, texture feature. Propounded classifier contains bayes classifiers, learning vector quantization and multilayer perception. On the basis of Eigen face recognition, Propounded technique used automated classification of WBC.

ISBN 978-1-7281-6202-7

Sheikh H et al. [11] ascertained the important cellular constituent of blood cells are Erythrocyte and leukocyte and blood platelet. Their size volume and characteristics are different. Here in their work, they propounded methods to identify blood using artificial neural network. They used incidental blood slides viewing 100X magnification and store them for analysis. Features extraction was done through wavelet transformation algorithm. Artificial neural network stores extracted features. The stored wavelets are used as input for background propagation. The cell images sizes are reduced by application of wavelet pyramids and wavelet coefficients are sorted and that parameters are used in ANN. Final recognition rate in this experiment was 88.9%. Lin QM et al. [12] presented a WBCs segmentation approach which is consist of three steps First Mathematical morphology is used to apply initial label for each pixel of blood image for thresholding. The algorithm is based on learning process. Knowledge of blood is applied for labeling on large regional information. Better segmented result was obtained by a fuzzy patch relaxation algorithm. In the shape correction process, they scan the images and to identify false and real region in real images. They named two regions as Cytoplasm region and leukocyte region by regularity detection method. Hough transformation was applied to detect a circular edge points in rough region of cell's image. In this method they successfully applied to 25 slides and segmentation done and identify RBCs and able to mark the irregular shapes to avoid the mistakes. Bikhet SF et al. [13] presents a method to recognize different types of normal WBCs by set of pre-treatment and segmentation algorithms with average success rate of 91%. Misclassification is generated by predefined stored values of brightness and area, which conclude misclassification of cell. Histogram entropy classification treated by an iterative threshold selection was used to get threshold value. In Cell type extracted edge has been splited by cell, nucleus and cytoplasm, ratio of cell area to nucleus area, ratio of nucleus area to circumference, nucleus circularity, cytoplasm average color, nuclei quantity, cell circularity, & zero crossing. The confusion matrix applied to compare the likeness and dislikeness. Approached this test to correct the classification of cell was 90%. NilufarS et al. [14] chose histogram of features computed on segmented cell and its nucleus and two-dimensional joint histogram classification problem and also investigated feasibility of three-dimensional histogram for their application. They propounded a method of combination of generative and discriminative classifier for variable length of sequences applied to detect remote protein. And also propounded automatic segmentation algorithm based joint histogram and propounded Bhattacharya kernel. For cell boundary they propounded Dirichlet GVT snake. Bhattacharya coefficient which is gives cross-section of two statistical samples to efficiently measure the relative closeness of two joint histogram and that calculation is defined by mathematical function. This increases accuracy of image classification significantly. Krishna Kumar Jha et al. [15] is actually based on leukemia detection is based on pre-processing which is based on segmentation through mutual information-based hybrid model-based extraction & classification. The Fuzzy C means algorithm and Active contour model is transformed to MI based hybrid model used for segmentation.Then propounded CSCA based DCNN classifier for classifications receive statistical and LDP features of segmented image.

A. challenges

Challenges erupted for Leukemia detection according to the FAB classification.

- A mathematical model was developed by many researcher but failure to reach proper result.
- Exhausting and tidy work of pathologist for manual predication of WBCs from single BSIs.
- For the segmentation, some technique is failed to differentiate WBCs from BSIs.
- The identification process depends the image quality like image resolution, image size. Those reason are responsible for some detection technique.

Pre-existing leukemia detection techniques contains huge amount of limitation, which work as the motivation, this paper designed and developed using FAB based SVM and NBC for WBC detection. Propounded method tried to overcome limitations of the pre-existing leukemia detection techniques.

III. PROPOUNDED METHED FOR LEUKEMIA DETECTION

Paper proposes the FAB - based SVM and NBC to identify leukemia. Block Diagram (Fig.2) illustrates the detection mechanism for leukemia using the propounded method. Firstly, the BSIs, which behave as the input, to upgrade the nature of input image pre-processing are applied. A set of patterns for the simplification was generated by segmentation of input image. By employing feature extraction of segmented image, feature like area, color, texture and shape are extracted. The method performs FAB - based SVM and NBC for the classified of the number of single blood cells.

A. Image pre-processing

An image is expressed by the mathematical function g (m, n) where m and n are two co-ordinates. Database of single cell is considered. For better classification results, images are being pre-processed. In this paper for image pre-processing was performed throughread image, resize image, grey scale, Remove noise (Denoise), Morphology (smoothing edges).

B. Propounded hybrid model for segmentation

After the pre-processing, the single BSIs are going for segmentation via the hybrid model. Cytoplasm & Nucleus are segmented. This research has blossomed the FAB classification-based hybrid model is shown as fig.3. The propounded segmentation needs to perform through the following sequence. Initially, the pre-treatment image passed through the hybrid segmentation, which combine the propounded LBG algorithm, hysteresis and watershed thresholding and adaptive canny. Propounded FAB

ISBN 978-1-7281-6202-7

classification-based hybrid segmentation model is illustrated by following diagram.

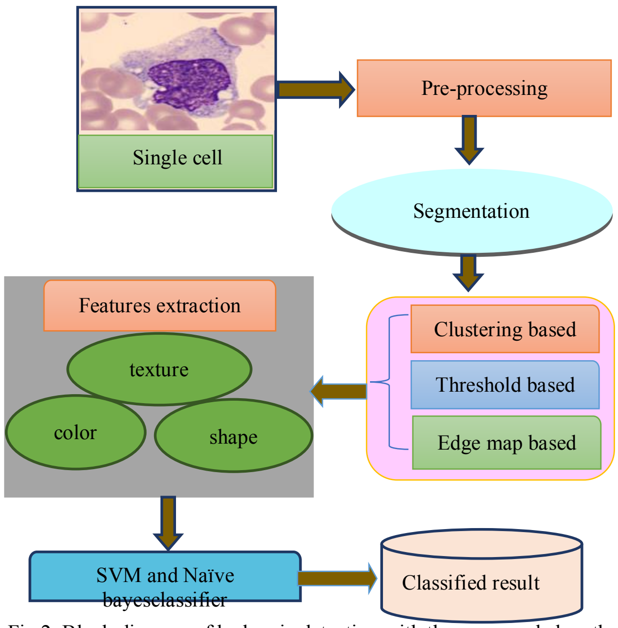


Fig 2. Block diagram of leukemia detection with the propounded method.

1) Segment through threshold based model

Segmentationcan be achieved by separating intensity of different region of image pixel by applying threshold value. This paper propounds the threshold-based technique which is improved the segmentation outcome on next level. Hysteresis is the lingering of an effect. The propounded Hysteresis thresholding can perform following in this case,

a. (Pixels > maximum threshold) => object and
b. (Pixels <minimum threshold) => background.
c. Minimum threshold <Pixels <maximum thresholds =>object ⇔ they area adjacent to other object pixels.

Watershed thresholding method is an effective mathematical tool for image segmentation. There are various techniques to implement the watershed method. Based on closeness of two region-of-interest with each other, segmentation of image is defined by his technique.

2) Segment through clustering based algorithm

The linde-buzo-grey (LBG) algorithm is utilized on the dataset to split the data point; it is close to k-means clustering. Propounded algorithm used for designing of codebook, which represent the minimum error and distortion. LBG choose the codebook randomly then chosen codebook set as a centroid and their vector through nearest distance. After that it find the new centroid of each group to get new codebook, if it not successful to capture the new codebook it repeats iteratively from initial code book, the most important reason to call iterative method. It has set of input vector is given equation

$$M = \{P_i \in S^t \mid i = 1,2,\ldots\ldots.,n\} \quad (1)$$

And generate the output vector is given equation

$$N = \{Q_j \in S^t \mid j = 1,2,\ldots\ldots.,m\} \quad (2)$$

And the user described $\ll n$, propounded clustering method need maximum iterations to find optimal solutions. Compute the distortion and analysis successfully of our propounded method.

3)Propounded method for edge map-based model

The propounded adaptive canny used to filter to estimate the noise and find the appropriate edge of WBCs. Our propounded technique not only can detect the edge map but it is self-adaptability. Many authors propounded different edge detect algorithms which are sobel, zero cross, log, Roberts, prewitt, traditional canny etc, but our propounded method is more efficient than others method. It is extract edge map effectively and have strong anti-noise ability. Compute the gradient $\psi_\sigma(\sigma,\tau)$ and $\psi_\tau(\sigma,\tau)$ and the equation is given

$$\psi_\sigma(\sigma,\tau) = \frac{1}{2}[\mu(\sigma+1,\tau) - \mu(\sigma-1,\tau)] \quad (3)$$

$$\psi_\tau(\sigma,\tau) = \frac{1}{2}[\mu(\sigma,\tau+1) - \mu(\sigma,\tau-1)] \quad (4)$$

After calculating the gradient perform the equation (5)

$$\mu(\sigma,\tau) = \frac{1}{\text{Ж}} \sum_{\omega=-1}^{1} \sum_{\upsilon=-1}^{1} [\mu(\sigma+\omega,\tau+\upsilon)\, \text{Ґ}(\sigma+\omega,\tau+\upsilon)] \quad (5)$$

Where, $\text{Ж} = \sum_{\omega=-1}^{1} \sum_{\upsilon=-1}^{1} \text{Ґ}(\sigma+\omega,\tau+\upsilon)$ (6)

is the parameter which represent the smoothness of the detected WBCs.

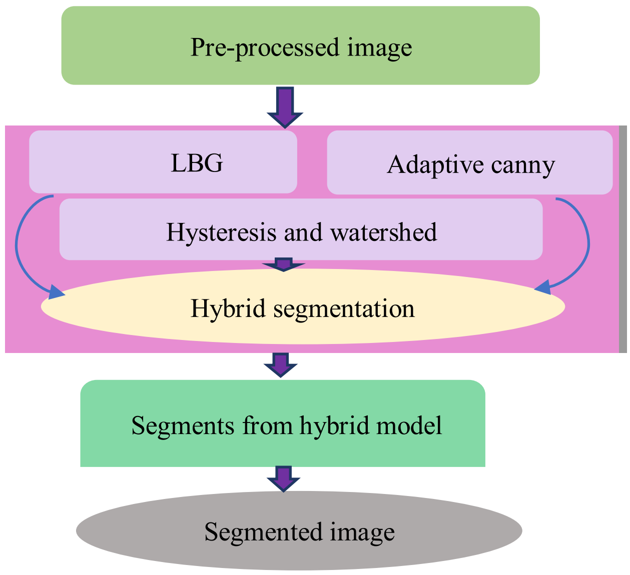


Fig 3. Work flow of Segmentation by propounded hybrid model

C. Features extraction process through segmented samples

A huge amount of unnecessary data is redefined into many relevant characteristics with minimum dimensions by new method. Different types of WBCs were recognized by feature extraction methods. The features extracted with better accuracy. Near closeness was considered while feature extraction, which lead to separation of normal and abnormal cell. The spatial variation of intensities of pixel of image provide texture features. Texture feature need for separation of

 ISBN 978-1-7281-6202-7

same size cells, it also need to separate normal and abnormal cells. Finally total non-zero values among total pixel provides classified information about cell.

D. Propounded SVM and Naïve-Bayes for classification

SVMs are supervised learning module used for classification algorithm along with associated learning paradigm that synthesizes data utilized for regression and classification analysis. It can simply solve Linear A non-linear problems. It creates a hyperplane that segments data into classes. A boundary line needs to create to distinguish the hyperplane between two classes of data. The function of line is

$$l = \delta\beta + \gamma \tag{7}$$

Hyperplane equation is

$$\alpha.\beta + \gamma = 0 \tag{8}$$

compute the hyperplane defines the hypothesis functions

$$h_{(\beta i)} = \begin{cases} +1 \; if \; \alpha.\beta + \gamma \geq 0 \\ -1 \; if \; \alpha.\beta + \gamma < 0 \end{cases} \tag{9}$$

Where, +1 is denote the above classified hyperplane and -1 represent the below of the hyperplane.

Naive Bayes or probabilistic classifiers that utilize Bayes' Theorem. It assumes that availability of specific feature in a class. This model outperforms even in highly sophisticated classification. Because they undergo through relatively short training. This method can solve Classification problems.

$$\text{Ƥ}(\text{T}|\text{Ɋ}) = \frac{\text{Ƥ}(\text{Ɋ}|\text{T})\text{Ƥ}(\text{T})}{\text{Ƥ}(\text{Ɋ})} \tag{10}$$

Where,

$\text{Ƥ}(\text{T}|\text{Ɋ}) = \text{Ƥ}(\text{Ɋ}_1|\text{T}) \times \text{Ƥ}(\text{Ɋ}_2|\text{T}) \times \ldots. \times \text{Ƥ}(\text{Ɋ}_n|\text{T}) \times \text{Ƥ}(\text{T})(11)$

here, $\text{Ƥ}(\text{T}|\text{Ɋ})$ is refers to posterior probability, where T is target class and Ɋ is attributes, $\text{Ƥ}(\text{T})$ refers to prior probability and $\text{Ƥ}(\text{Ɋ})$ denote the prior predictor probability.

The classification process is given bellow

Initial new data = $(\text{Ɋ}) = (\text{Ɋ}_1, \text{Ɋ}_2, \ldots., \text{Ɋ}_n)(12)$

Class T is member of $\{\text{T}_1, \text{T}_2, \ldots., \text{T}_m\}(13)$

IV. RESULT AND DISCUSSION

This section presents the experimental outcomes achieved by the propounded FAB classification based SVM and NBC for the leukemia detection. For the experimentation, the research considers the images from the dataset master and cellavison database [3].

a. Dataset description

The single cell blood smear samples for the experimentation of the propounded classifier are taken from dataset master and cellavison database. The dataset master database has the large collection of trimmed section of blast cells and healthy with different type of WBCs. Contrary, cellavion database with single cell blood smear samples is high resolutionwith small dimension.

b. Evaluation metrics

The performance of the propounded classifier is calculated on metrics, like True Positive Rate (TPR), True Negative Rate (TNR) and accuracy.

By using propounded method, detect the five type of WBCs:

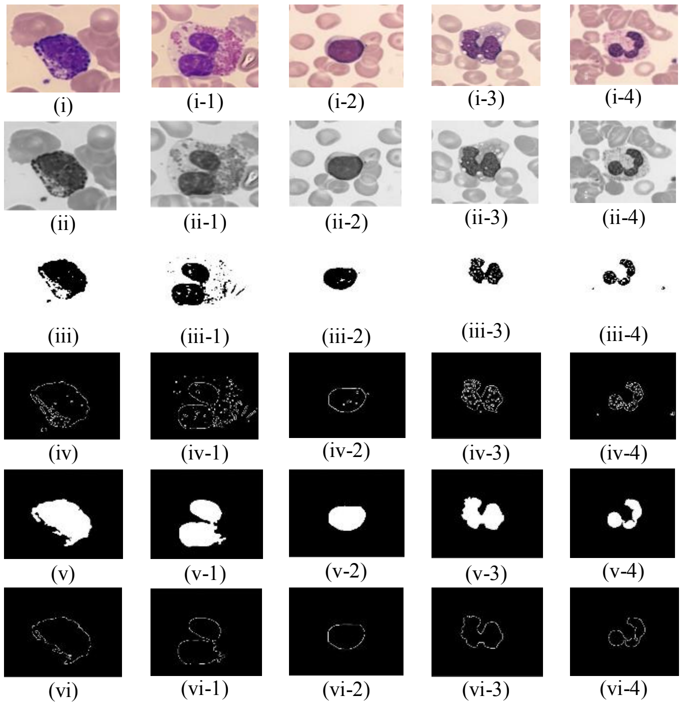


Fig 4: experimental result achieved by propounded segmentation model five type of single BSIs (i-vi).

c. Experimental results

Fig.4represents the experimental results of the propounded method during segmentation task. The results of five type of WBCs sample show the detection processes through our propounded method. Fig. 4i is input image sample, Fig. 4ii show the grey scale image, and Fig. 4iii is illustrate binary image with propounded threshold method from grey scale image sample and propounded LBG used to image achieve the best results. Fig 4iv depicts the propounded edge detection method, and Fig. 4v represent hole fill image which successfully help the edge detect, and finally Fig. 4vi depicts the propounded adaptive canny detect the final edge for leukemia.A propounded model can able to separate all five types of WBCs accurately which is shown in Fig. 4.

d. Accuracy comparisons of our propounded method

The existing methods, such as Sayed [16], HSVM [17], SVM and random forest [18] are compared with mixed database of dataset master and cellavision, the propounded classifier method to prove the effectiveness of the classification accuracy.

Comparative analysis						
Methods	Basophil (%)	Eosinophil (%)	Lymphocyte (%)	Monocyte (%)	Neutrophil (%)	Classification accuracy (%)
New Method	99.3	99.2	99.6	98.4	98.8	99.06
SVM & random forest[18]	100	70	74.8	85.3	97.1	92.8
HSVM [17]	43.8	0	66.8	0	7.5	76.3
Sayed [16]	53	63	85.0	39.0	50.8	76.8

e. *Performance analysis*

The propounded method compares with other methods and shows the detection process by means of performance analysis. The comparisons results achieve much better accuracy almost then some other method, and our propounded classifier perform well. Training, testing and validation-based performance are given the fig. 5, it presents the flow of Training, testing and validation-based performance. Mean square error (MSE) is 0.0407, which is indicate that our classifier performance and neural network is perform thewhole analysis of large data. The performance is carried out the different training percentage and the k-fold value. The five type of WBCs is segment successfully and detect the proper edge map through our propounded method.

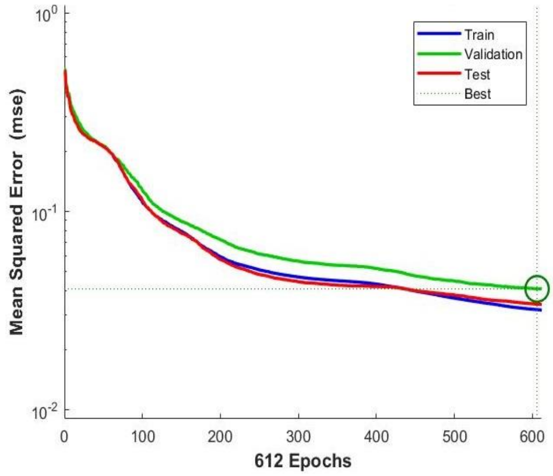


Fig 5: presented performance on training data, testing data and best validation set.

f. *Performance based on training data, testing data and validation*

Here, fig. 5 depicts the performance on training, testing data and validation set rest of the work. We generate the graph with help of the neural-network-toolbox where the validation set is utilized to find the goal and optimized the best model to decode the problem. The validation set approach the model of estimate rate. Training dataset fit the model and it is utilized to provide unbiased evaluations of a final model fit on the training dataset.

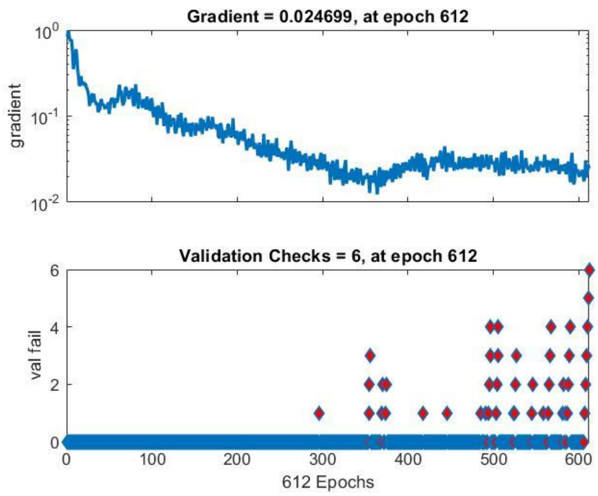


Fig 6: presented performance of gradient and validation check based on data.

g. *Performance curve of gradient and validation based on data.*

Fig. 6 represents the performance of the gradient and validation with epoch. Here, the gradient value noted as 0.2469 at epoch 612 and validation is 6 at epoch 612 where epoch is calculated based on no. of times every vector are used to change the weight.

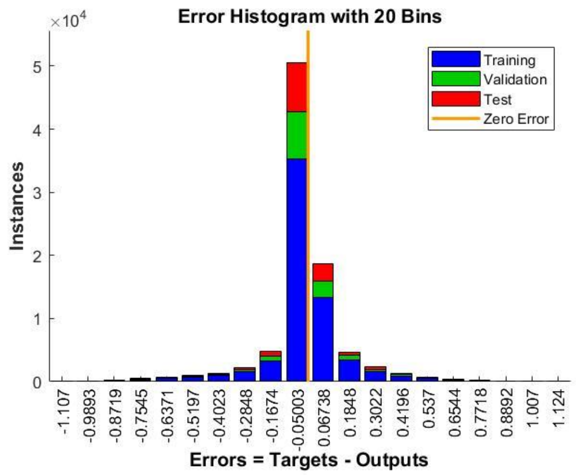


Fig 7: presented performance of error based onbetween target and actual output.

h. *Performane of error between target and actual output*

This section represent the error of our particular model where error form NN ranges strating from -1.107 to 1.124. The total error divided into 20 bins. After calculation of the errors, each bin having width of 0.1115. Fig.7 contain vartical bar which represent no. of samples from dataset.

V. CONCLUSION

This paper is concentrated to detect and identify leukemia. After performing number of experiments, it justified that edge of each type of WBCs is characterized using propound algorithms as it is shown edge detection being implemented by means of adaptive canny whereas clustering is being performed by LBG algorithm. Thresholding was successfully implemented by hysteresis and watershed thresholding. After performing number of experiments on number of sample images Success rate is almost 99.06 %. But still huge improvement required. So, we used hybrid model of SVM and Naïve Bayes for classification. So,system will become independent by means of use of neural network. Of course, our classification method needs more improvement. Currently we had used less data and images which is the basic reason for getting best result, while Neural Network has better performance with large data. Therefore, how to improve the classification and prediction of disease based on our method is another direction of our study in the future.

FUNDING

This research activity is financially supported by R&D project, Memo No: 148(Sanc.)/ST/P/S&T/6G-13/2018 of Science & Technology and Biotechnology Dept, Government of West Bengal, India.

ACKNOWLEDGEMENT

The authorstake this opportunity to extend their heartliest thanks to Dr. Rupama Chaudhuri, DNB (Pathology), CMO, FWC, Mr. Biplab Kanti Das, Asst. Prof., and Dr. Himadri Mandal, Asst. Prof. for physical verification of outcome by visual verification and valuable feedback provided time to time for research activity.

REFERENCES

[1] Krishna Kumar Jha, Himadri Sekhar Dutta, "Nucleus and cytoplasm–based segmentation and actor-critic neural network for acute lymphocytic leukaemia detection in single cell blood smear images," Medical & Biological Engineering & Computing,58(1):171-186. doi: 10.1007/s11517-019-02071-1, 2020.

[2] R. Bhattacharjee, L.M. Saini, "Robust technique for the detection of acute lymphoblastic leukemia," Proceedings of the IEEE Power, Communication and Information Technology Conference (PCITC), pp. 657–662, 2015.

[3] Cellavision Inc(2019). http://www.cellavision.com/.

[4] H.Q. Huang, X.Z. Fang, J. Shi, J. Hu, "Abnormal localization of immature precursors (ALIP) detection for early prediction of acute myelocytic leukemia (AML) relapse," Journal of Medical&Biological Eng. &Computing 52 (2), pp. 121–129, 2014.

[5] M. Sezgin, "Survey over image thresholding techniques and quantitative performance evaluation," Journal Electron. Imaging 13 (1),pp. 146–168, 2004.

[6] I. Cseke, "A Fast Segmentation Scheme for White Blood Cell Images", Proceedings. 11th IAPR International Conference Pattern Recognition, Conference C: Image, Speech and Signal Analysis, vol III, pp. 530-533, 1992.

[7] P.L. Rosin, "Measuring shape: ellipticity, rectangularity, and triangularity", Proceedings. 15th International Conference on Pattern Recognition, vol. 1, pp. 952 – 955, 2000.

[8] P. Yampri, C. Pintavirooj, S. Daochai, S. Teartulakarn, "White blood cell classification based on the combination of eigen cell and parametric feature detection," Proceedings of the 1st IEEE conference on industrial electronics and applications (ICIEA 06), pp 1–4, 2006.

[9] J. Schurmann, "Pattern Classification: A Unified View of Statistical and Neural Approaches," New York: Wiley, 1996.

[10] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Mach. Learn., vol. 46, no. 1–3, pp. 389–422, Jan. 2002.

[11] Sheikh H, Zhu B, Tzanakou EM, "Blood cell identification using neural networks," Proceedings of the IEEE 22nd annual northeast bioengineering conference, pp 119–120, 1996.

[12] QM Lin, YY Deng, "An accurate segmentation method for white blood cell images," IEEE International Symposium on Biomedical Imaging 2002, pp. 245–248, 2002.

[13] Bikhet SF, Darwish AM, Tolba HA, Shaheen SI, "Segmentation and classification of white blood cells," IEEE International Conference Acoust Speech Signal Process 4, pp. 2259–2261, 2000.

[14] Nilufar S, Ray N, Zhang H, "Automatic blood cell classification based on joint histogram-based feature and Bhattacharya Kernel," Proceedings of the 42nd Asilomar conference on signals, systems and computers (ASILOMAR 08), pp. 1915–1918, 2008.

[15] Krishna Kumar Jha, Himadri Sekhar Dutta, "Mutual Information based hybrid model and deep learning for Acute Lymphocytic Leukemia detection in single cell blood smear images," Computer Method and Program in Biomedicine,vol. 18, pp. 3-43, 2019.

[16] Rezatofighi SH, Soltanian-Zadeh H, "Automatic recognition of five types of white blood cells in peripheral blood," Computerized Medical Imaging and Graphics, vol. 35(4), pp. 333–343, 2011.

[17] Tai WL, Hu RM, Hsiao HCW, Chen RM, Tsai JJP, "Blood cell image classification based on Hierarchical SVM," IEEE International Symposiumon Multimedia, ISM 2011, pp. 129–136, 2011.

[18] Jianwei Zhao, Minshu Zhang, Zhenghua Zhou, Jianjun Chu,Feilong Cao1, "Automatic detection and classification of leukocytes using convolutional neural networks,"Medical & Biological Engineering & Computing Doi 10.1007/s11517-016-1590-x, 2016.