Classification of White Blood Cell Images Using Probabilistic Neural Networks

GUMMADIDALA ASHA[1], ADAPA DEEPTHI[2], BANDARU SOWMYA[3], AJIMEERA KAVERI BAI[4], DR. PULAGAM AMMI REDDY[5]

[1, 2, 3, 4] Student, Department Of Electronics and Communication Engineering, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India

[5] Professor, Department Of Electronics and Communication Engineering, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India

Abstract- Numerous diseases can be diagnosed based on the number of cells for each class of white blood cells in the blood. Therefore, extracting information about that is considered very important for haematologists. The density of WBC's in our blood stream provides a glimpse into state of our immune system. There are five types of WBC's. They are Neutrophil, Eosinophil, Basophil, Monocyte and Lymphocyte. This project presents a better way to classify the White Blood Cells. The proposed method is implemented with the help of K Means clustering with Probabilistic Neural Networks (PNN) classifier .The proposed K Means clustering with PNN classifier gives better results.

Indexed Terms- White Blood Cells (WBCs), Neutrophil, Eosinophil, Basophil, Monocyte, Lymphocyte, K Means clustering, Probabilistic Neural Networks (PNN).

I. INTRODUCTION

White blood cells account for only about 1% of your blood, but their impact is big. White blood cells are also called leukocytes. They protect you against illness and disease.

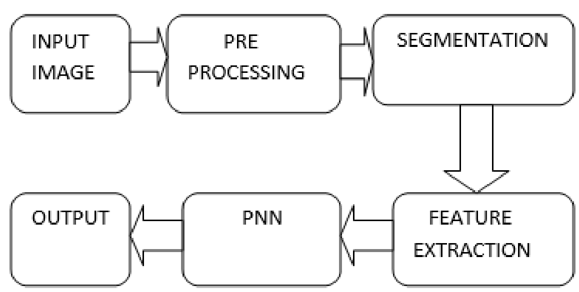
Think of white blood cells as your immunity cells. In a sense, they are always at war. They flow through your bloodstream to fight viruses, bacteria, and other foreign invaders that threaten your health. When your body is in distress and a particular area is under attack, white blood cells rush in to help destroy the harmful substance and prevent illness.

The main job of white blood cells, or leukocytes, is to fight infection. There are several types of white blood cells and each has its own role in fighting bacterial, viral, fungal, and parasitic infections.

- Help heal wounds not only by fighting infection but also by ingesting matter, such as dead cells, tissue debris, and old red blood cells.
- Protect you from foreign bodies that enter the blood stream, such as allergens.
- Are involved in the protection against mutated cells, such as cancer.

The existing method is based on Artificial Neural Networks (ANNs) and having a limited accuracy, precision, sensitivity and specificity of 80% in its performance evaluation metrics.

II. PROCESSING FLOW



III. PROPOSED METHOD

With the technological advances in medical field, the need for faster and more accurate analysis tools becomes essential for better patient's diagnosis. In this work, the image recognition problem of white blood cells (WBC) is investigated. Five types of white blood cells are classified using a feed forward back propagation neural network. After segmentation of blood cells that are obtained from microscopic images, the most 16 significant features of these cells are fed as inputs to the neural network. Half of the 100 of the WBC sub-images that are found after segmentation are used to train the neural network, while the other half is used for test. The results found are promising with classification accuracy being 96%.

So here for the detection purpose we are using of K-Means Clustering with Probabilistic Neural Networks (PNN).

3.1 INPUT IMAGE

For the classification of white blood cell images here we use blood smear images that are taken under a digital microscope with 400X magnification specifications by means of image pre-processing techniques such as image acquisition, pre-processing, image segmentation, edge detection feature extraction.

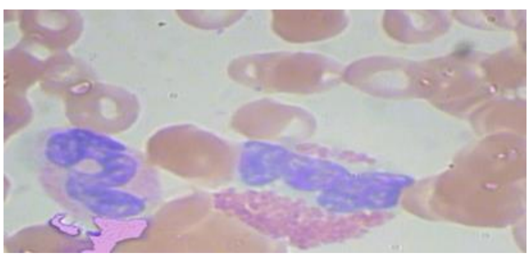


Fig 3.1 Input image

3.2 PRE-PROCESSING

Pre-processing helps in improvement of the image data by suppressing unwanted distortions.

In the microscopic images which are given as input image consists of Gaussian noise and shot noise. Gaussian low pass filter is used to remove the noise in the image and smooth the image for further process which is segmentation.

Gaussian function is represented as

$$i(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$$

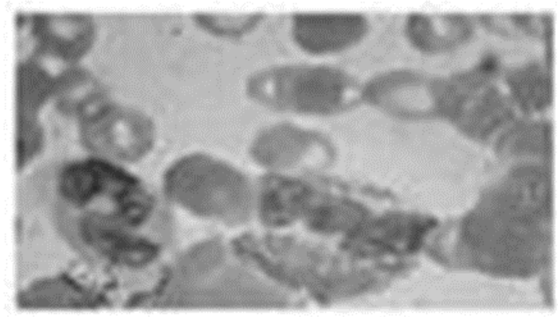


Fig 3.2 Pre-processed image

3.3 SEGMENTATION

The division of an image into meaningful structures is called as image segmentation. It is often an essential step in image analysis, object representation, visualization, and many other image processing tasks.

There are various techniques for image segmentation but in this project K-means clustering is used for image segmentation.

3.3.1 K-MEANS ALGORITHM

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different.

We use it here to denote techniques that are primarily used in exploratory data analysis of high-dimensional measurement patterns. This K-Means clustering is done through K-Means algorithm.

$$L = \sum_{i=1}^{m} \sum_{k=1}^{k} w_{ik} \left\| x^i - r_k \right\|^2$$

Where w_{ik}=1 for data point x^i if it belongs to cluster *k*; otherwise, w_{ik}=0. Also, r_k is the centroid of x^i's cluster.

K-means algorithm is an iterative algorithm that tries to partition the dataset into *K* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the

inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. The way k-means algorithm works is as follows:
Step 1: Specify the number of clusters (k).
Step 2: Allocate objects to clusters.
Step 3: Compute cluster mean.
Step 4: Allocate each observation to the closest cluster centre.
Step 5: Repeat steps 3 and 4 until the solution converges.

The output from the segmented block is as follows:

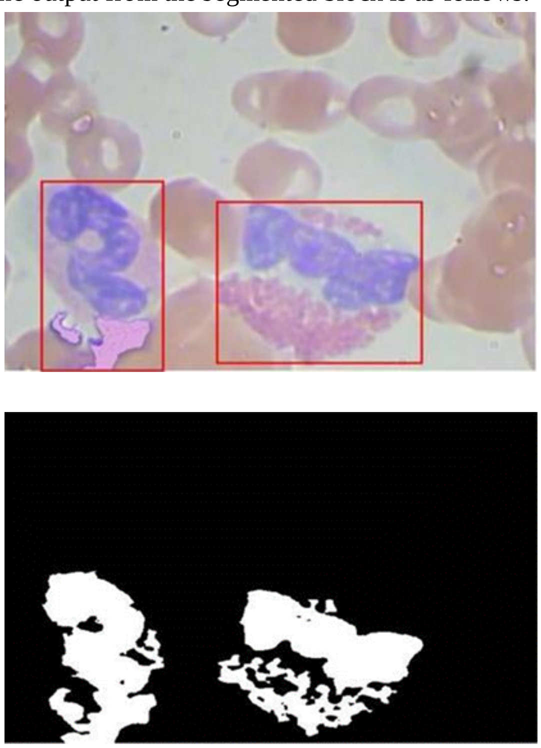


Fig 3.3 Segmented image

3.4 FEATURE EXTRACTION

Feature extraction is done based on intensity, shape and texture.

Texture is a measure of the variation of the intensity of a surface, quantifying properties such as smoothness, coarseness and regularity.

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix.

3.4.1 GLCM

Grey Level Co-Occurrence Matrix (GLCM) is one of the best known texture Analysis methods. GLCM estimates image properties related to second order statistics. Each entry (i , j) in GLCM corresponds to the number of occurrences of the pair of grey levels i and j which are a distance d apart in the original image.

The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.

Algorithm:
Step1: Quantize the image data.
Step2: Create the GLCM.
Step3: Make the GLCM symmetric:
i) Make a transposed copy of the GLCM
ii) Add this copy to the GLCM itself.
Step4: Normalize the GLCM.
Following texture features are calculated using GLCM.

$$Energy = \sum_{i,j=0}^{N-1} (p_{ij})^2$$

$$Contrast = \sum_{i,j=0}^{N-1} p_{ij}(i-j)^2$$

$$Entropy = \sum_{i,j=0}^{N-1} -\ln(p_{ij})(p_{ij})$$

$$Homogenity = \sum_{i,j=0}^{N-1} \frac{p_{ij}}{1+(i-j)^2}$$

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-M)(j-M)}{\sigma^2}$$

For the proper analysis on feature extraction and Gray Level Co-occurrence Matrix we followed the references from [12] to [16].

3.5 PNN

Probabilistic Neural Networks is a supervised machine learning algorithm which can be used for both classification and regression challenges.

The objective of the Probabilistic Neural Networks algorithm is to find a hyper plane in an N-dimensional space (N-the number of features) that distinctly classifies the data points.

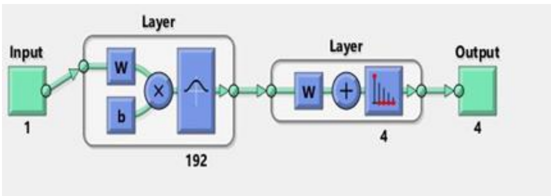


Fig 3.4 PNN Layers

A probabilistic neural network (PNN) has 3 layers of nodes. The figure shown in the results displays the architecture for a PNN that recognizes K = 2 classes, but it can be extended to any number K of classes. The input layer (on the left) contains N nodes: one for each of the N input features of a feature vector. These are fan-out nodes that branch at each feature input node to all nodes in the hidden (or middle) layer so that each hidden node receives the complete input feature vector x. The hidden nodes are collected into groups: one group for each of the K classes as shown in the figure. Each hidden node in the group for Class k corresponds to a Gaussian function centered on its associated feature vector in the k class (there is a Gaussian for each exemplar feature vector). All of the Gaussians in a class group feed their functional values to the same output layer node for that class, so there are K output nodes.

PNN is a Bayes– Parzen classifier. Bayes Parzen classifier could be broken up into a large number of simple processes implemented in a multi layer neural network. The univariate case of PDF was proposed by Parzen. The multi variate PDF estimator is given as follows

$$g(x_1, x_2 \ldots x_P) = \frac{1}{n_{\sigma_1 \sigma_2 \cdots \sigma_n}} \sum_{i=1}^{n} w \left(\frac{x_1 - x_{1,i}}{\sigma_1}, \frac{x_2 - x_{2,i}}{\sigma_2}, \ldots \ldots . \frac{x_p - x_{P,i}}{\sigma p} \right) \sigma_1{}' \sigma_2, \cdots \sigma_p$$

Where $\sigma_1{}'\sigma_2, \cdots \sigma_p$ are the smoothing parameters representing standard deviation around the mean of p random variables $x_{1,} x_2, \cdots x_p$. W is a weighted function to be selected with specific characteristics.

$$y(x) = \frac{1}{(2\Pi)^{P/2} n_{\sigma p}} \sum_{i=1}^{n} \exp\left(\frac{\|x - x_i\|^2}{2\sigma^2} \right)$$

Where x is the vector of random variable and x_i is the ith training vector. Weighting function is as follows

$$w(d) = \frac{1}{1 + d^2}$$

As sample size n increases the Parzen's PDF estimator asymptotically approaches the true underlying density function.

For proper analysis on Probabilistic Neural Networks we followed references from [2] to [8].

The feature extracted image is then processed through PNN and the WBC gets separated and that separated image is used in order to classify the type and the results are as follows.

Fig 3.5 Classified WBC

IV. EXPERIMENTAL RESULTS

Classifier Name	Features	Classifier Accuracy
ANN	30	89.3
MLP-BP	42	91.2
KNN	47	93.0
PNN	50	94.0

Table 1: Classification Performance Comparison

Nearly 369 images are used for the classification of WBCs in our project. In those 70% of images are used for training , 15% images for validation and remaining 15% for testing The input data set is divided into sub groups. Each group consist of 47 images in those 23 are considered as first category and 24 are considered as second category. In those 5 images are used for validating fake images. And 6 for validating real images.

The accuracy achieved with PNN classifier for the classification of WBC images is 94%. Remaining 6% being the false classification.

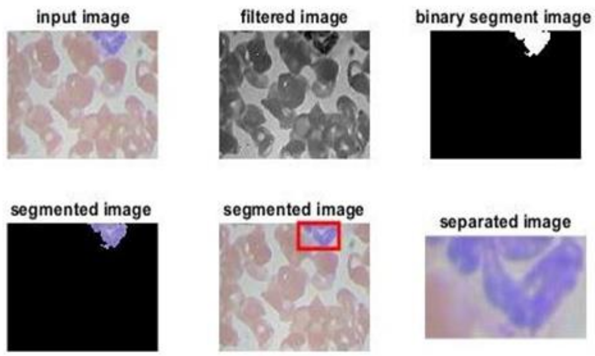


Fig 4.1 WBC image after Preprocessing and segmentation

CONCLUSION

The probabilistic Neural Networks trained by Feed Forward Neural Networks have been used to classify 5 types' white blood cells. Several features are used as inputs to the probabilistic neural networks. The choice of feature and the classifier plays a significant role in accuracy. Segmentation of images is done with an accuracy of 97% using K Means clustering algorithm and PNN classifier classifies with an accuracy of 90. Finally WBCs are classified with an accuracy of 90-95%. By improving the preprocessing and segmentation standards we can achieve 98-99% accuracy in classifying WBCs.

REFERENCES

[1] L. P., J, Gartner, L. Hiatt, "Concise Histology", Elsevier, 2011, ISBN: 978-0-7020-3114-4.

[2] S. Mu-Chun, Chun-Yen Cheng, and Pa-Chun Wang, "A Neural-Network-Based Approach to White Blood Cell Classification", The Scientific World Journal, Volume 2014, Article ID 796371, 9 pages.

[3] M. Nørgaard, O. Ravn, N. K. Poulsen, and L. K. Hansen, "Neural Networks for Modelling and Control of Dynamic Systems", Springer-Verlag, London, 2000.

[4] T. S. Mohammed and Nidal Ibrahim al-Tataie, "Artificial Neural Network as a Decision- Makers for Stereo Matching", STF- International Journal on Computing Vol. 1, No. 3, pp (89 – 94), August 2011.

[5] M. S. Al-Ani, Thabit Sultan Mohammed, and Karim M. Aljebory "Speaker Identification: A Hybrid Approach Using Neural Networks and Wavelet Transform", Science publications, Journal of Computer Science, 3 (5): 404-409, 2007.

[6] R. Simutis, D. Dilijonas, L. Basting, and J. Fraiman, "A Flexible Neural Network for ATM Cash Demand Forecasting", 6th WSEAS Int. Conf. on Computational Intelligence, Man-Machine Systems and Cybernetics, Tenerife, Spain, December 14-16, 2007.

[7] N., Ganesan, "Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data" International Journal of Computer Applications. 2010, Vol.1, No 26, pp.76-85.

[8] G. Ongun, Halici U., Leblebicioglu K., AtalayV., Beksac M., and Beksac S., "Feature Extraction and Classification of Blood Cells for an Automated Differential Blood Count System," in Proceedings of the International Joint Conference on Neural Networks, USA, pp.2461-2466, 2001.

[9] S. Hsieh, Wang Z., Cheng P., Lee I., Hsieh S., and Lai F., "Leukemia Cancer Classification Based on Support Vector Machine," in Proceedings of the 8th IEEE International Conference on Industrial Informatics, Japan, pp.819-824, 2010.

[10] A. Abdul Nasir, Mashor M.Y., and Hassan,R.," Classification of Acute Leukaemia Cells using Multilayer Perceptron and Simplified Fuzzy ARTMAP Neural Networks", The International Arab Journal of Information Technology, Vol. 10, No. 4, July 2013.

[11] M. Z. Othman,., and Ali, Alaa, B.," Segmentation and Feature Extraction of Lymphocytes WBC using Microscopic Images", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,Vol. 3 Issue 12, December-2014.

[12] K. Amandeep, and Rajneesh, R., "Content-based Image Retrieval: Feature Extraction Techniques and Applications", Mandi Gobindgarh:

International Conference on Recent Advances and Future Trends in Information Technology,2012.

[13] A. Gebejes, and R. Huertas ,"Texture Characterization Based on Grey-Level Co-occurrence Matrix", Granada: Conference of Informatics and Management Sciences, 2013.

[14] K. Vijay, and Priyanka, G. ," Importance of Statistical Measures in Digital Image Processing" ,Hyderabad: International Journal of Emerging Technology and Advanced Engineering, 2012.

[15] O. Suhail, Manal, K.," Apply Multi-Layer Perceptrons Neural Network for Off-line signature verification and recognition" . International Journal of Computer Science.Vol. 8, Issue 6. Bethlehem.

[16] A. Gebejes, A. and Huertas, R.," Texture Characterization Based on Grey-Level Co-occurrence Matrix", Granada: Conference of Informatics and Management Sciences, 2013.

[17] S. Rashmi, and S. Mandar, " Textural Feature Based Image Classification Using Artificial Neural Network", Advances in Computing, Communication and Control. Communications in Computer and Information Science, vol 125. Springer, 2011.