Detection of the White Blood Cell using Clustering Approach and classification using Multi-layer Architecture

Anshu Jaswal[1],Astha Gautam[2]

M.Tech (Scholar), Assistant Professor

Department of computer science Engineering, LRIET College, Solan

Abstract- Leukaemia Detection structure analyses the microscopic picture and overcome these problems. It removes the necessary parts of pictures and direct applies some methods. K-mean collecting is used only WBC (WHITE BLOOD CELL) detection. In this thesis we define a system for medical data processing that mostly uses Sift (scale invariant feature transformation), best solution identifies that the Ant Colony Optimization Techniques and classification has been done (Back Propagation Neural Network). Its offers the consequences achieved by dispensation dissimilar data including databases of children with learning structure damage. Blood pictures of the good pixel quality are obtained. The noise is removed from the image using filter. Categorize the white blood cell. Solidity essential to be measured for clear image. Roundness checks whether the shape is circular or not .To remove the features from the indulgence picture. Basically find out the features of the nucleus of myelocyte & lymphocytes. Feature Extraction by SIFT is the processing of converting the image into data. After feature extraction we can identify the best solution with the help of ant Colony Optimization Technique. Input picture the BPNN classifier that test image into either infested or not infection. Here ALL-IDB proposed by Donida Labati. The major part of this work is to segment the white blood cells for leukaemia detection. An evaluate the performance parameters alike false acceptance rate, false refusal rate, accuracy and compare the white blood cell in manual and auto count.

Keywords: *Leukaemia Detection, White Blood Cell, Feature Extraction (SIFT), K-mean Algorithm and Classification.*

I. INTRODUCTION

Medical imaging has turn into one of the most important formation & explanation methods in ecology & medicine above the previous decade. This time has perceived an incredible expansion of novel, prevailing apparatuses for identifying, packing, conducting, study & exhibiting medical pictures. This has led to enormous growth in the application of digital image dispensation techniques [1] for cracking medical difficulties.

Data Mining also called as information discovery in databanks is procedure of extracting potentially helpful information from fresh data. A software apparatus can examination large quantity of data & automatically statement attractive patterns without requiring human intervention [1]. Other knowledge finding technologies are arithmetical study, OLAP, information Visualization, and Ad-hoc questions. Different these machineries, data mining do not require a human to ask specific questions.

Leukemia is the cancer of the blood. It starts in the bone marrow [3], it is the area where blood cells are complete. When you have leukaemia, the bone marrow begin to make a lot of abnormal white blood cubicles, called leukaemia cells. They don't do the effort of usual white blood cells. They grow faster than normal cells, and they don't break increasing while they should. Above time, leukaemia cells can mass out the normal blood cells. This cans chief to serious problems such as anaemia, bleeding, & infectivity. Leukaemia cells can also spread to the lymph nodes or other organs and origin bulge or pain [2].

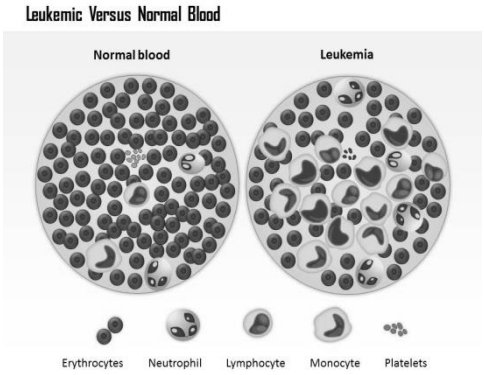


Figure 1 leukemia

There are numerous different types of leukaemia. In specific, leukaemia is together by how fast it gets poorer and what kind of white blood cell it change.

A.Types of Leukemia

- Acute lymphoblastic leukaemia.
- Chronic lymphocytic leukaemia.
- Acute myelogenous leukaemia.
- Chronic myelogenous leukaemia.Indications may be contingent on what type of leukaemia you have, although general symptoms include [4]:
- A new lump or swollen gland in your collar, below your arm, or in your projection.
- Recurrent nosebleeds, flow from the gums or rectum, additional frequent staining, or extremely heavy menstrual bleeding.
- Frequent fevers.
- Night sweats.
- Bone aching.
- Inexplicable appetite defeat or recent weight defeat.
- Feeling tired a lot without a known cause.
- Swelling & pain on the left sideways of the belly.

B. White Blood Cell

White blood cells are bigger in size than the red blood cells. The attentiveness & masterpiece of the white blood cells give some imperative information which helps us to find out many diseases. White blood cells can be considered in to 5 types: Neutrophil, Basophil, Eosinophil, Lymphocyte and Monocyte which are shown in the Figure 2 These cells fight against diseases and protect our body [5].

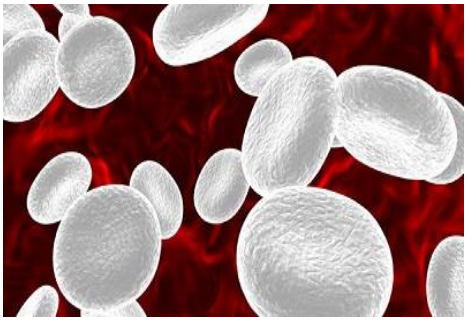


Figure 2 White Blood Cells

C. Types of White Blood Cells

1. Neutrophil
2. Eosinophil
3. Basophil
4. Lymphocyte

II. RELATED WORK

R. Hassan et.al,2012 [6] presented the study on blasts categorising in acute leukaemia into two chief forms which was acute myelogenous leukaemia and acute lymphocytic leukaemia by using k-NN. 12 main features that represent size, colour-based and shape were extracted from acute leukaemia blood images. **Liviu Badea et.al,2012 [7]**They were integrating the biggest publicly available gene expression data sets of leukaemia & usual haematopoiesis with the aim of uncovering the main gene modules involved in normal haematopoiesis as well as in the various leukaemia subtypes. **Dong Ling Tong et.al,2014 [8]**aimed to sought a systems biology approach to suppose gene supervisory networks of leukaemia related markers. Consequences demonstrated the efficacy of a systems ecosystem approach to make simpler composite genetic interactions without losing significant biological data of the genes. Key indicators that inter-connected other leukaemia associated markers were identified. **Chaitali Raje et.al,2014 [9]**proposed system was on microscopic images to detect Leukaemia. The first and fast identification of Leukaemia greatly aids in provided that the suitable treatment. Personalize segmentation is complete using Statistical restriction such as mean, model deviation which separates white blood cells since other blood components i.e. erythrocytes & platelets. Geometrical structures such as area; perimeter of the white blood cell nucleus is investigated for diagnostic prediction of Leukaemia. **Mashiat Fatma et.al,2014[10]**was proposing a technique for correct and quick classification of leukaemia pictures & cataloguing them into their personally types. For this, different features are extracted from the input images and then based on these geographies a data set for the input images were created. This data set is then utilized as input data to a neural network for exercise purposes. This neural network had designed & created to categorize the pictures according to their equivalent leukaemia type.

III. ISSUES IN DETECTION OF THE WBC CELLS

The literature on the leukocyte segmentation, it is noticed that large number of methods are only working on the extraction of nucleus but there are very few methods available which are extracting the cytoplasm and even with less accurately. The main reason behind the less accuracy in the cytoplasm extraction is that most of the researchers are using the grey level colour for the extraction of cytoplasm which is not easily separable from the other colours. It is noticed in the literatures that different approaches are used for the white blood cells detection. Some have used KNN approach, threshold techniques, EM algorithm, Fuzzy rules, watershed transform, GVF model, trained neural network, Fuzzy c-mean clustering, computer morph metric system and many more. From the literature studied, it has been observed that there are many ways we can make a better system for the identification of leukaemia from the microscopic blood image. None of the researchers has

used the K-mean clustering for the segmentation of the white blood cells from the microscopic blood image. In this thesis K-mean clustering approach has been used on the clean microscopic blood image followed by image cleaning and the extraction of the nucleus and cytoplasm with a good accuracy. In our proposed work, classify the data using BPNN (Back Propagation Neural Network) to find the accuracy. In the previous work, accuracy has not reached to greater extent, so man motive will be to enhance the accuracy set.

IV. PROPOSED TECHNIQUE

A. Feature Extraction Using Sift (Scale Invariant Feature Extraction)

The SIFT approach, for picture highlight era, takes [11] a picture and changes it into an "expansive gathering of neighbourhood highlight vectors" (From "Article Recognition from Local Scale-Invariant Features", David G. Lowe). Each of these highlight vectors is invariant to any scaling, revolution or interpretation of the picture. This methodology offers numerous highlights with neuron reactions in primate vision. To help the extraction of these highlights the SIFT calculation applies a 4 stage separating methodology:

- Scale-Space Extreme Detection

This phase of the separating endeavours to recognize those areas and scales that are identifiable from diverse perspectives of the same item. This can be proficiently accomplished utilizing a "scale space" capacity. Further it has been demonstrated under sensible suppositions it must be in light of the Gaussian capacity. The scale space is characterized by the capacity:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

Where * is the difficulty administrator, G(x, y, σ) is a variable-scale Gaussian and I(x, y) is the data picture.

Different strategies can then be utilized to distinguish stable key point areas in the scale-space. Distinction of Gaussians is one such procedure, finding scale-space extreme, D(x, y, σ) by registering the contrast between two pictures, one with scale k times the other. D(x, y, σ) is then given by:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

To recognize the nearby maxima and minima of D(x, y, σ) every point is contrasted and its 8 neighbours at the same scale, and its 9 neighbours here and there one scale. On the off chance that this worth is the base or greatest of every one of these focuses then this point is an extreme.

- Key point Localisation [11]

This stage endeavours to take out more focuses from the rundown of key points by discovering those that have low difference or are inadequately restricted on an edge. This is accomplished by computing the Laplacian, see mathworld.wolfram.com/Laplacian.html for subtle elements, esteem for every key point found in stage 1. The area of extremism, z, is given by:

$$Z=- \ 〚(d\hat{}2 \ D\hat{}(-1))/〚dx〛\hat{}2 \ 〛\hat{} \ dD/dx$$

On the off chance that the capacity esteem at z is underneath a limit esteem then this point is prohibited. This uproots extreme with low difference. To dispose of extreme taking into account poor localisation it is noticed that in these cases there is a huge guideline shape over the edge yet a little ebb and flow in the opposite headline in the deference of Gaussian capacity.

On the off chance that this distinction is underneath the proportion of biggest to littlest eigenvector, from the 2x2 Hessian lattice at the area and size of the key point, the key point is rejected.

- Introduction Assignment

This step intends to relegate a predictable introduction to the key points taking into account nearby picture properties. The key point descriptor, depicted underneath, can then be spoken to with respect to this introduction, attaining to invariance to pivot. The methodology taken to discover an introduction is

Utilize the key points scale to choose the Gaussian smoothed picture L, from above

Figure angle size, m

Figure introduction, θ.

Structure an introduction histogram from inclination introductions of test focuses

Find the most noteworthy top in the histogram. Utilize this top and whatever other nearby crest inside 80% of the stature of this top to make a key point with that introduction

A few focuses will be relegated various introductions.

Fit a parabola to the 3 histogram values nearest to every top to interject the tops position.

- Key point Descriptor

The nearby inclination information, utilized above, is additionally used to make key point descriptors. The slope data is pivoted to line up with the introduction of the key point

and after that weighted by a Gaussian with change of 1.5 * key point scale. This information is then used to make an arrangement of histograms over a window focused on the key point.

Key point descriptors normally utilizes an arrangement of 16 histograms, adjusted in a 4x4 lattice, each with 8 introduction containers, one for each of the attitude compass bearings and one for each of the mid-purposes of these headings [12].

B. Ant Colony Optimization Technique

- Swarm intelligence revisions the cooperative presentation of unsophisticated agents that interact locally through their situation [13].
- It is motivated by social insects, such as ants and termites, or other animal societies, such as fish schools & bird flocks.
- Although every separate has only limited capabilities, the complete swarm exhibitions complex complete behaviour. Therefore, the intellectual behaviour can be seen as an emergent distinguishing of the swarm.
- When converging on ant colonies, it can be professed that ants communicate only in an indirect manner through their atmosphere by dropping a element called pheromone.
- Paths with higher pheromone levels will more likely be preferred & thus reinforced, while the pheromone strength of pathways that are not chosen is decreased by evaporation.
- This form of indirect declaration is known as stigmergy, & offers the ant colony shortest-path finding capabilities.
- ACO employs imitation ants that cooperate to find good solutions for discrete optimization difficulties. These software managers mimic the foraging performance of their biological complements in finding the shortest-path to the food source.

C. Back Propagation Neural Network

The Back Propagation neuron system is artificial neural net based on error back propagation procedure. The Back Propagation Neural Network imitation consists of an enter layer, some hidden layers and an output layer. Each connection linking neurons has a characteristic weighting value. In training the system, the knobs in the BP neural network find input information as of exterior basis, & then go by to hidden layer which is an interior data processing layer & is responsible [14] for the information conversion, & then the nodes in the output layer source the required output matter. Behind that, the anti-propagation of mistake is dejected by distinct the real output with wanted output. Every weight is reconsider & back propagated layer by layer from productivity layer to hidden layer & input layer. This procedure will be continued until the output fault of system is reduced to an acceptable level or the determined time of education is realized. The dispensation consequences of information are exported by output layers to the outside.

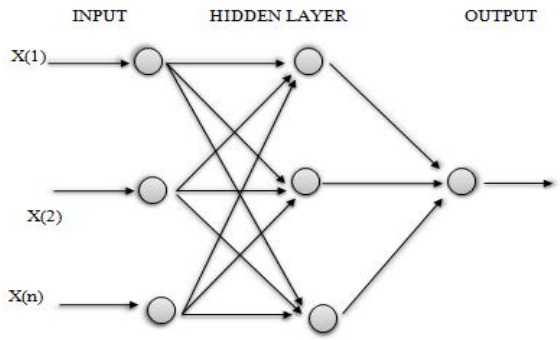


Figure 3. Back Propagation Neural Network

BP neural network consists of a lot of neurons that are efficient in a form of 3 layers: input, hidden & output. The neurons are connected by weights W y In training the system with a specified architecture, the back propagation approach, finds a single best set weight rate by minimization of appropriate error purpose. In a multi-layer feed forward neural network, the processing elements are in order the layers & only the rudiments in adjoining layers are connected. It has a smallest of three layers of basics (i.e., input layer, the middle or hidden layer, & the productivity layer). The designation "back propagation" (BP) creates as of the fact that scheming are passed feed familiar from the input layer to the output layer, following which considered errors are broadcast back in other way to change the weights to obtain a better performance. BP method is an expansion of the smallest mean square method that can be used to train multi-layer systems. The 3-layered free forward neural network is displayed in Fig. 3 which is comprised by input layer, hidden layer and output layer.

V. SIMULATION MODEL

Step I. First we upload sample the dataset images form the hospital dataset.

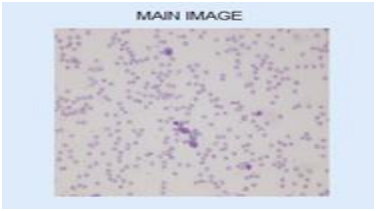


Figure 4. Main Image

Step II: To find the white blood cell, red blood cell and background image using zack algorithm.

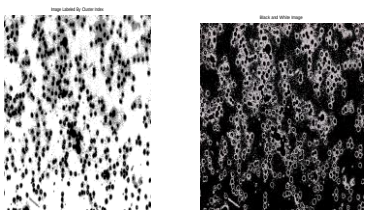


Figure 5. Cluster Index image and Segmented Image

Step III: Apply k-mean algorithm for white blood cell. This algorithm used for divide the cell in two forms and identifies the affected and non-affected data in this image.

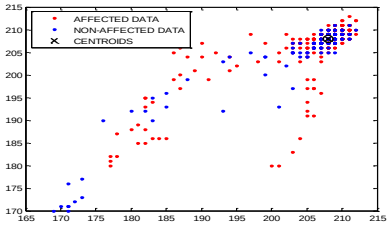


Figure 6. K-mean Clustering Image

Step IV: To count the white blood cell.

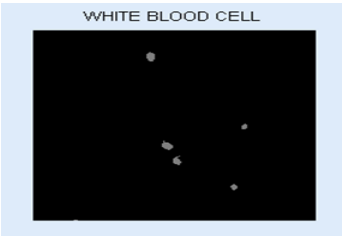


Figure 7. WBC cell Image

Step V: To extract the feature through sift algorithm (identify the unique properties). Each of these highlight vectors is invariant to any scaling, revolution or interpretation of the picture. This methodology offers numerous highlights with neuron reactions in primate vision.

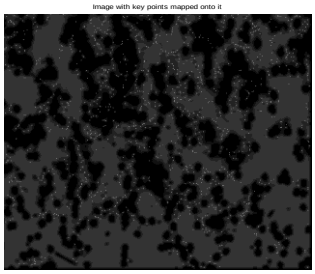


Figure 8 Feature Extracted using SIFT

Step VI: After Uniqueness found then to optimize the features using Swarm Intelligence Field (ACO) Algorithm. Swarm intelligence revisions the cooperative presentation of unsophisticated agents that interact locally through their situation [13]. It is motivated by social insects, such as ants and termites, or other animal societies, such as fish schools & bird flocks

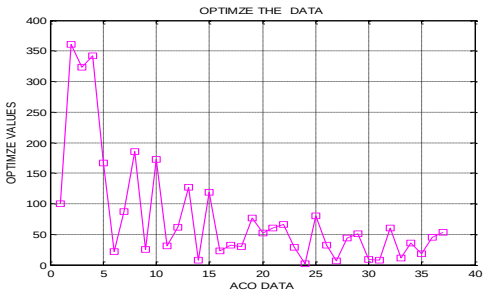
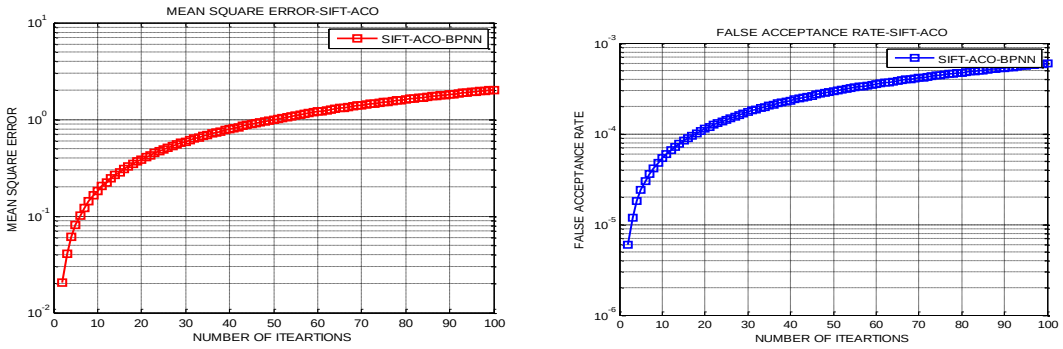


Figure 9 ACO Optimization

Step VII: To classify the features using Back Propagation Neural Network. BP neural network consists of a lot of neurons that are efficient in a form of 3 layers: input, hidden & output. The neurons are connected by weights W y In training the system with a specified architecture, the back propagation approach, finds a single best set weight rate by minimization of appropriate error purpose. In a multi-layer feed forward neural network, the processing elements are in order the layers & only the rudiments in adjoining layers are connected.

VI. RESULT AND DISCUSSIONS

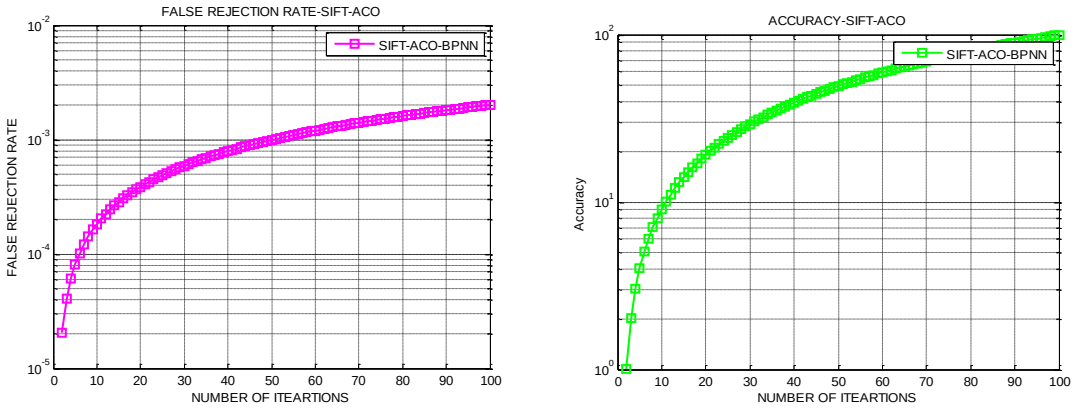
In this section defined that the result and discussion of the performance parameters i.e Mean Square Error, False Acceptance Rate and False Rejection Rate.



(i) (ii)

Figure 10. (i) Mean Square Error Rate and (ii) False Acceptance Rate

Figure 10(i) defines; the mean square error rate means average of the training and testing module error find using BPNN+SIFT and Ant Colony Optimization algorithm. Figure shows 10(ii) ; false acceptance rate means positive data find using BPNN+SIFT and Ant Colony Optimization algorithm.



(i) (ii)

Figure 11.(i) False rejection Rate and 12(ii) Accuracy

Figure shows, the false rejection rate means negative data collect using Artificial Intelligence technique used for multiple layer data transfer and ACO algorithm.

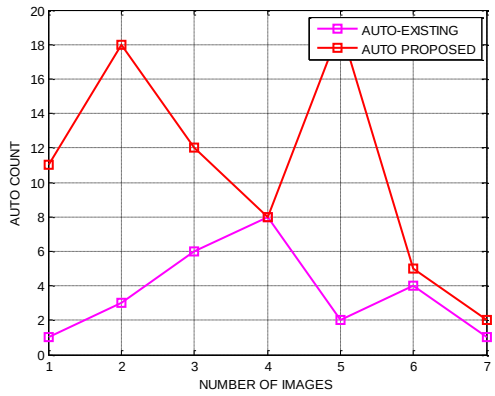
Above shows; Accuracy means correct the system design.

Table no: 1 Existing Work

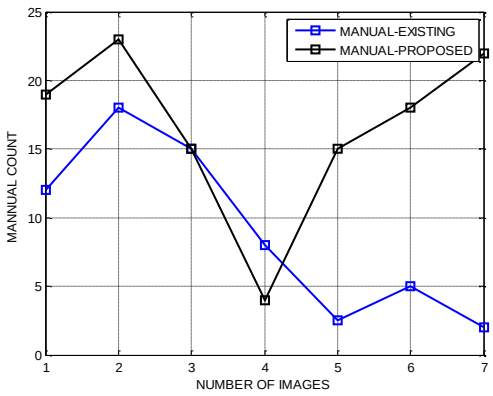
Image no	Manual Count Exiting work	Auto count Existing Work	Accuracy in Existing Work
Img 1	12	11	91
Img 2	18	18	80
Img 3	15	12	100
Img 4	8	8	79
Img 5	2.5	19	87
Img 6	5	5	82
Img 7	2	2	92

Table no: 2 Proposed Work

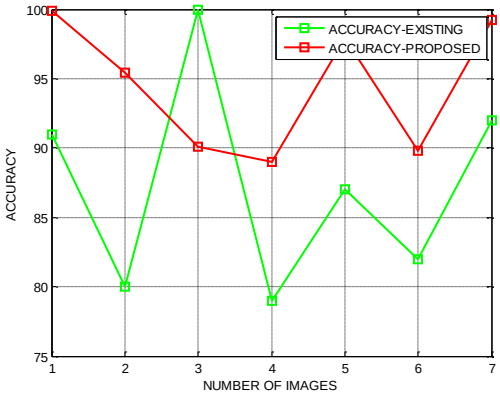
Image no	Manual Count Proposed work	Auto count Proposed Work	Accuracy in Proposed Work
Img 1	19	1	99.9
Img 2	23	3	95.4
Img 3	15	6	90.1
Img 4	4	8	89.0
Img 5	15	2	97.9
Img 6	18	4	89.8
Img 7	22	1	99.2



(i)



(ii)



(iii)

Figure 12. Comparison between auto count, manual count and accuracy(proposed and Existing)

Fig 12.(i) Comparison Between auto count(previous and proposed),(ii) Comparison between Manual Count(previous and proposed) and (iii) Comparison between Accuracy (existing and Proposed) Above figure 12(i) shows that the auto count proposed work and previous work found maximum manual count is 8and minimum manual count is 1. Above figure shows 12(ii) that the manual count proposed work and existing work found maximum manual count is 23 and minimum manual count is 4. Above figure shows 12(iii) that accuracy find the proposed work and previous work . Maximum proposed value is 99.89 and minimum value is 89.

Table no: 3 Comparisons between False Acceptance Rate, False Rejection Rate and Accuracy in 50 images

IMAGE NO	REJECTION RATE	ACCEPTANCE RATE	ACCURACY
IMG 15	0.00057125	3.0965e-05	99.93
IMG 16	0.00057109	0.00019251	99.92
IMG 17	0.00057118	0.00037341	99.936
IMG 18	0.00057092	0.00036506	99.9064
IMG 19	0.00057113	0.00015482	99.927
IMG 20	0.00057104	0.00023864	99.91
IMG 21	0.00057125	3.6174e-05	99.93
IMG 22	0.00057105	0.00023619	99.91
IMG 23	0.00059113	0.0015627	99.92
IMG 24	0.00057091	0.00037087	99.9058
IMG 25	0.00057126	1.8075e-05	99.0491
IMG 26	0.00057125	4.782e-05	99.9381
IMG 27	0.0001764	0.00040252	99.783
IMG 28	0.0017606s	0.00019258	99.8047
IMG 29	0.0017603	0.0005035	99.773

IMG30	0.0017604	0.00033115	99.7908
IMG31	0.0017607	0.00010447	99.8135

IMG 32	0.0017605	0.00021142	99.8028
IMG 33	0.0017601	0.0006155	99.76
IMG 34	0.0017605	0.00025787	99.79
IMG 35	0.0017604	0.0040368	99.78
IMG 36	0.0026263	0.0012522	99.612
IMG 37	0.0026264	0.0012451	99.6129
IMG 38	0.002626	0.0013523	99.5822
IMG 39	0.0026273	0.00032763	99.7045
IMG 40	0.002626	0.0015985	99.5776
IMG 41	0.0026263	0.0012942	99.60
IMG 42	0.0026264	0.0011969	99.6177
IMG 43	0.0026275	0.0011399	99.7259
IMG 44	0.0026271	0.005263	99.6847
IMG 45	0.0026272	0.00037099	99.7002
IMG 46	0.0026201	0.00147	99.5902
IMG 47	0.0026275	0.0001213	99.7251
IMG 48	0.0026275	0.0001067	99.7266
IMG 49	0.0026277	0.00016748	99.7205
IMG 50	0.002626	0.00097775	99.6396
IMG 51	0.0026272	0.00041459	99.6957

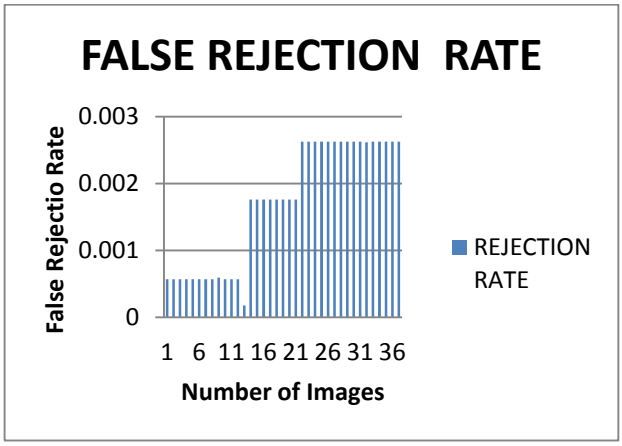


Figure 13. False Rejection Rate in 50 images

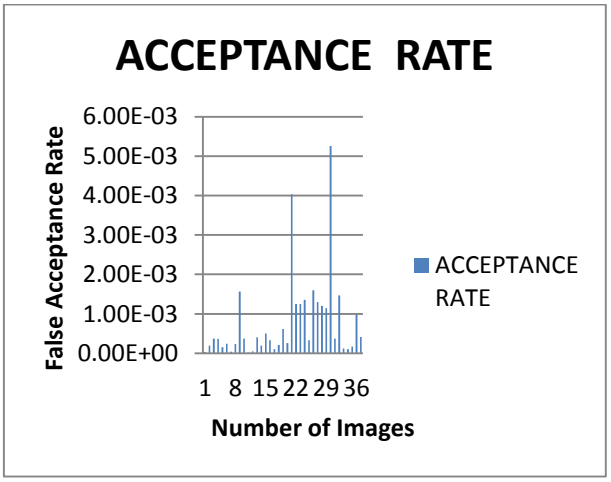


Figure 14. False Acceptance Rate in 50 Images

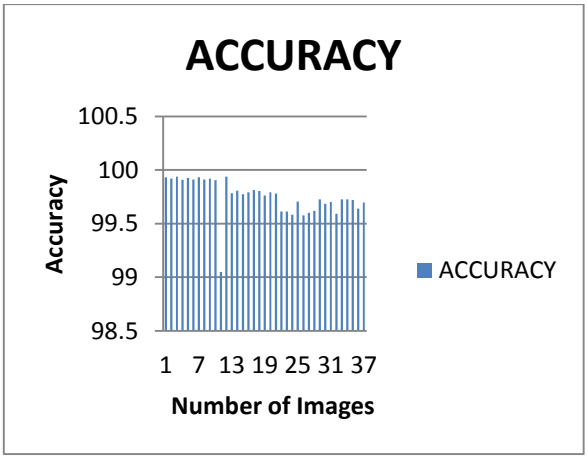


Figure 15. Accuracy in 50 Images

VII. CONCLUSION

Main centre of this paper is to research an automated system which can detect the leukaemia from the microscopic image to improve the accuracy and reduce the time to detect than the manual approach .So many lives can be save by using the proposed approach of leukaemia detect. The feature extraction image using Principle Component Analysis feature transforms. To detect the unique properties and in the form of matrix i.e key points in different positions. After extraction we apply the problem solving algorithm using Ant Colony Algorithm. The main part of this work is to segment the white blood cell for leukaemia detect. The training module using Back propagation Neural Network for classification purpose. First, epochs means how many numbers of epochs to complete the training module, time consider and performance. The accuracy gets proposed system is 99.89% .We can also use the purposed system to find out the percentage of leukaemia infection in microscope image. We hope this approach will be beneficial for today's fast life and early detection of leukaemia without any need of costly tests and with a better accuracy.

REFERENCES

[1]. Tang, Wenlong, Hongbao Cao, and Yu-Ping Wang. "Subtyping of Leukemia with Gene Expression Analysis Using Compressive Sensing Method."Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on. IEEE, 2011.

[2]. Mohapatra, Subrajeet, et al. "Fuzzy based blood image segmentation for automated leukemia detection." Devices and Communications (ICDeCom), 2011 International Conference on. IEEE, 2011.

[3]. Jiang, Yong, et al. "Extractives of Rumex restrain the proliferation of leukemia cell THP-1." IT in Medicine and Education (ITME), 2011 International Symposium on. Vol. 2. IEEE, 2011.

[4]. Halim, NH Abd, et al. "Nucleus segmentation technique for acute leukemia."Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on. IEEE, 2011.

[5]. Madhloom, Hayan T., Sameem Abdul Kareem, and HanyAriffin. "A Robust Feature Extraction and Selection Method for the Recognition of Lymphocytes versus Acute Lymphoblastic Leukemia." Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on. IEEE, 2012.

[6]. Supardi, N. Z., et al. "Classification of blasts in acute leukemia blood samples using k-nearest neighbour." Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on. IEEE, 2012.

[7]. Badea, Liviu. "Unsupervised analysis of leukemia and normal hematopoiesis by joint clustering of gene expression data." Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on. IEEE, 2012.

[8]. Tong, Dong Ling, and Graham R. Ball. "Exploration of leukemia gene regulatory networks using a systems biology approach." Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on. IEEE, 2014.

[9]. Raje, Chaitali, and JyotiRangole. "Detection of Leukemia in microscopic images using image processing." Communications and Signal Processing (ICCSP), 2014 International Conference on. IEEE, 2014.

[10]. Fatma, Mashiat, and Jaibir Sharma. "Identification and classification of acute leukemia using neural network." Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014 International Conference on. IEEE, 2014.

[11]. Scotti F. Robust segmentation and measurements techniques of white cells inblood microscope images. In: Daponte P, Linnenbrink T, editors. Proceedingsof the IEEE instrumentation and measurement technology conference, 24–27April. Sorrento, Italy: IEEE Publisher; 2006. p. 43–8.

[12]. Piuri V, Scotti F. Morphological classification of blood leucocytes by microscope images. In: Proceedings of the IEEE international conference on computational intelligence for measurement systems and applications, 14–16 July. Boston,MA, USA: IEEE Publisher; 2004. p. 103–8.

[13]. Halim NHA, Mashor MY, Hassan R. Automatic blasts counting for acute leukemia based on blood samples. Int J Res Rev Comput Sci 2011;2(August(4)).

[14]. Ning, Yuan, et al. "Comparison of different BP neural network models for short-term load forecasting." Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on. Vol. 3. IEEE, 2010.