

Neural Network Classification of White Blood Cell using Microscopic Images

Mazin Z. Othman

Electrical and Computer Eng. Dept.,
College of Eng., Dhofar University,
Salalah, Oman

Thabit S. Mohammed

Electrical and Computer Eng. Dept.,
College of Eng., Dhofar University,
Salalah, Oman

Alaa B. Ali

Computer Science,
Bayan University,
Erbil, Iraq

Abstract—With the technological advances in medical field, the need for faster and more accurate analysis tools becomes essential for better patients' 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%.

Keywords—White Blood Cell; Neural networks; Image analysis; Leukocytes; Lymphocyte; Feature extraction

I. INTRODUCTION

In the fields of haematology and infectious diseases, classifying different kinds of blood cells can be used as a tool in diagnosis. By counting certain cells' relative frequencies and comparing to what is normal, conclusions can be made about possible blood diseases. Blood consists of several elements which are white blood cell (WBCs), red blood cell (RBCs), platelets, and plasma. The quantity of blood cells plays important role to ensure the healthiness of a person.

Human blood contains five major types of WBC or what is referred to as leukocytes. The WBC types, which are illustrated in Figure 1, together with their typical relative frequencies are: *neutrophils*, *basophils*, *eosinophils*, *lymphocytes* and *monocytes*. In a human adult, the normal average number of WBC is about 7000/micro litre, which forms about 1% of the total blood cell in the body. The increase in the number of WBC in the body is referred to as leucocytosis, while decrease in the number of WBC is called leucopenia, with leucocytosis being the most likely to occur compared to leucopenia [1].

Due to the different morphological features of the white blood cells, manual classification of such cells is a cumbersome process, which is time-consuming and susceptible to human error as it is mostly related to the haematologists' experience. This fact actually emphasise a crucial need for fast and automated method for identifying the different blood cells.

Implementation techniques of automated differential blood cells counting systems are of two kinds [2]: One technique is based on the flow cytometry, while the other is based on

image processing. Image processing techniques are having the advantage over the flow cytometry based systems. In that the images of the blood samples can be saved and hence referred to for further verification in case some abnormal conditions were detected.

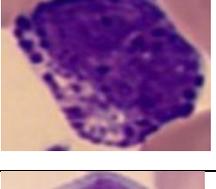
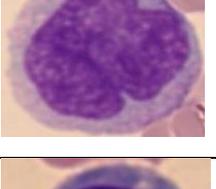
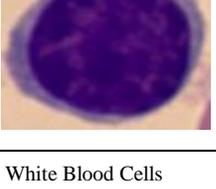
Cell Type	Relative Frequencies	Microscopic Image
NEUTROPHIL	(50-70)%	
EOSINOPHIL	(1-5)%	
BASOPHIL	(0-1)%	
MONOCYTE	(2-10)%	
LYMPHOCYTE	(20-45)%	

Fig. 1. Typical Images of Common White Blood Cells

In this work, the processing of microscopic images of blood cells using neural networks as an efficient decision maker for proper white blood cell type recognition is adopted. Neural networks have powerful features in analysing complex data, and among the wide and variant application areas of neural networks are the system identification and control [3], image recognition and decision making [4], speech and pattern recognition [5] as well as financial applications [6]. Artificial neural networks have also been successfully used in medical applications to diagnose several cancers [7].

As for the interest in this paper, which is the segmenting and classifying of blood cells microscopic images using neural network, a number of scientific researches have been published. Ongun et al. [8] have applied the multilayer perceptron network trained using conjugate gradient descent (CGD), linear vector quantisation (LVQ) and k-nearest neighbour classifier which produced 89.74%, 83.33% and 80.76% of accuracy, respectively.

Hsieh *et al.* [9] used the information gain technique based on support vector machine (SVM) for feature selection. Determining entropy and calculating the correlation in a training dataset, the usefulness of a feature was estimated while classifying the training data. The proposed technique was used to classify two types of leukaemia, which are acute lymphoblastic leukaemia (ALL) and acute myelogenous leukaemia (AML).

Abdul Nasir et al [10], proposed application of MLP and simplified fuzzy ARTMAP (SFAM) neural networks for classifying the individual WBC as lymphoblast, myeloblast and normal cell based on the extracted features from both acute lymphoblastic leukaemia (ALL) and acute myelogenous leukaemia (AML) blood samples. A total of 42 features (6 size, 24 shape, and 12 colour) were used and a classification accuracy of 93.82% is achieved.

In this work, the multilayer perceptron back-propagation MLP-BP neural network is used to classify the most known five types of WBC that have been segmented from blood smear microscopic images using the most distinguishing features. The adopted algorithmic comprises three stages. The first stage is image segmentation, the second stage is labelling that returns the number and location of each WBC, and the third stage is extracting descriptive features measured from the segmented cells.

II. PRE-PROCESSING AND SEGMENTATION

The algorithm adopted in this paper for image pre-processing and segmentation is basically proposed in [11]. The three main steps of this algorithm, which are segmentation, labelling, and feature extraction, are illustrated in Figure 2. The next image processing step is the WBC subtype recognition and this will be achieved by the use of neural networks.

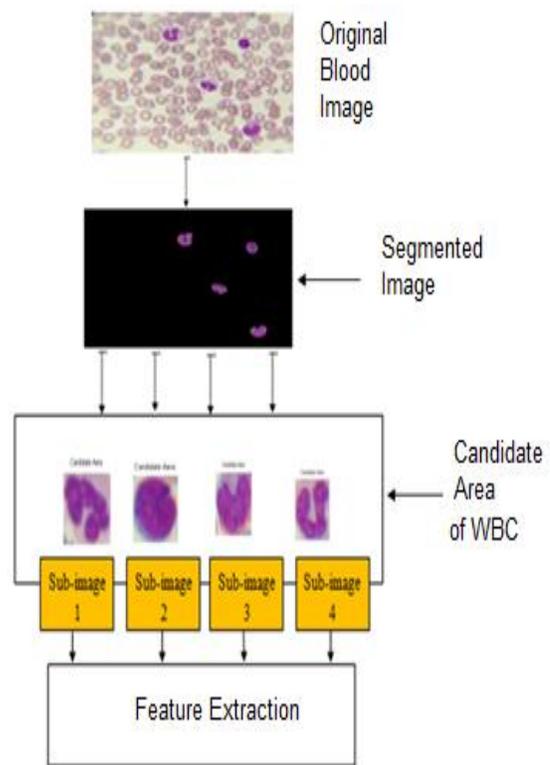


Fig. 2. The Pre-processing and Segmentation Procedure

III. FEATURE EXTRACTION OF WBC

The choice of features immensely affects the classifier performance. The features must characterize each WBC subtype and must be independent of each other for robust classification, better judgment and comparison. Indeed, an extensive work had been focused on determining different features that crucially distinguishing each type or groups of types of WBC. These features can be grouped into shape features, intensity features, and texture features.

A. Shape Features

There are many techniques of shape description and recognition. These techniques can be broadly categorised into two types: (1) boundary-based and (2) region-based [12]. The most successful representatives for these two categories are Fourier descriptor and moment invariants whereas moment invariants are to use region-based moments, which are invariant to transformations as the shape feature.

The regular moment invariants are presented by Hu [13] who derived a set of invariants using algebraic invariants.

In [11], it was found that compared with the set of invariant moments; the seventh moment invariant feature ϕ_7 has a noticeable effect on classification performance [11].

The two-dimensional seventh moment of a digitally sampled $M \times M$ image that has gray function $f(x, y)$, ($x, y = 0, \dots, M - 1$) is given as:

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (1)$$

where,

$$\eta_{pq} = \mu_{pq} / (\mu_{00})^\gamma \quad (2)$$

$$\gamma = \frac{p+q}{2} + 1 \quad (3)$$

In equations (2), μ_{pq} is calculated as follows:

$$\mu_{pq} = \sum_{x=1}^x \sum_{y=1}^y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (4)$$

where, (\bar{X}, \bar{Y}) are $\bar{X} = m_{10} / m_{00}$ and $\bar{Y} = m_{01} / m_{00}$ which are the centre of the image.

B. Intensity Features

These features are based only on the absolute value of the intensity measurements in the image. A histogram describes the occurrence relative frequency of the intensity values of the pixels in an image. The intensity features that will be considered are the first four central moments of this histogram: mean, standard deviation, skewness, and kurtosis.

For a grayscale image, the mean of the blood image is equal to the average brightness or intensity and it is given by:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{N} = \sum_{i=1}^N \frac{Xi}{N} \quad (5)$$

where, \bar{X} is the mean, N number of pixels, $X_1 \dots X_n$ are the grayscale image data.

The image variance, gives an estimate of the spread of pixel values around the image mean. The skewness measures the symmetry about the mean [14]. It is defined as:

$$Skewness = \frac{1}{N} \left(\frac{\sum_{i=1}^N (Xi - \bar{X})^3}{\sigma^3} \right) \quad (6)$$

The kurtosis (K) is a measure of whether the data are peaked or flattened, relative to a normal distribution and can be computed as: [15]

$$Kurtosis = \frac{1}{N} \left(\frac{\sum_{i=1}^N (Xi - \bar{X})^4}{\sigma^4} \right) - 3 \quad (7)$$

C. Textural Features

These features contain information about the spatial distribution of tonal variations within a band. Texture representation methods can be classified into three categories: (1) statistical techniques, (2) structural techniques, and (3) spectral techniques. Statistical techniques are most important for texture classification because these techniques result in computing texture properties [16].

1) The statistical techniques using gray level co-occurrence matrix (GLCM)

The identification of specific textures in an image is achieved primarily by modelling texture as a two-dimensional gray level variation. This two dimensional array is called gray level co-occurrence matrix (GLCM). The GLCM was originally proposed by R.M. Haralick, therefore features generated are known as Haralick features [17]. Co-occurrence matrix which is a tabulation of how often different combinations of pixel values occur in an image. Based on the co-occurrence matrices five texture features, namely, (1) contrast, (2) homogeneity, (3) entropy, (4) energy, and (5) correlation are calculated as in [16].

$$Contrast = \sum_{i,j} |i - j|^2 P(i, j) \quad (8)$$

where, i and j are the horizontal and vertical cell coordinates and p is the cell value.

$$H = \sum_{i,j} \frac{1}{1 - (i - j)^2} P(i, j) \quad (9)$$

$$Entropy = - \sum_{i,j=0}^{Ng-1} P(i, j) \log(p(i, j)) \quad \dots(10)$$

where, $P(i, j)$ is the (i, j) th entry of the normalised co-occurrence matrix, N_g is the number of gray levels of the blood image.

$$Energy = \sum_{i,j} P(i, j)^2 \quad \dots\dots(11)$$

$$Correlation = \frac{Conv(x_i, y_j)}{\sigma_i \sigma_j} = \frac{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i, j) P(i, j) - \mu_x \mu_y}{\sigma_i \sigma_j} \quad (12)$$

In Equation (12),

$$\mu_i = \sum_i i \sum_j P(i, j)$$

$$\mu_j = \sum_j j \sum_i P(i, j)$$

$$\sigma_i = \sum_i (i - \mu_i)^2 \sum_j P(i, j)$$

$$\sigma_j = \sum_i (j - \mu_j)^2 \sum_j P(i, j)$$

where, $\mu_x, \mu_y, \sigma_x,$ and σ_y are the means and standard deviations of the marginal probabilities $P_x(i)$ and $P_y(j)$ obtained by summing up the rows or the columns of matrix P_{ij} (co-occurrence matrix), respectively.

2) *The statistical techniques using colour moments*

The colour moments are the statistical moments of the probability distributions of colours which are the first order moment (mean) and the second order moment that are used in variance computation .

The mean of colour intensity of the RGB model is defined as:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \tag{13}$$

where, the i^{th} colour channel is defined at the j^{th} image pixel as p_{ij} and N is the number of pixels in the image [18].

The variance and the standard deviation are defined mathematically by equations (14) and (15):

$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \tag{14}$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{1/2} \tag{15}$$

where, f_{ij} is the value of the i^{th} colour components of the image pixel j , N is the number of features over all database, μ_i is the mean of the colour i .

I. NEURAL NETWORK CLASSIFICATION

The features that are considered significant to represent an image of white blood cells are extracted and accumulated in a vector, which we refer to as the features vector. Features vector is then transformed into a set of classes using neural networks as a technique to solve a WBC classification problem. This technique adopts a learning algorithm to identify a model that best fits the relationship between the feature set and class label of the input data. Therefore, a key objective of the learning algorithm is to build predictive model that accurately predict the class labels of previously unknown records.

The feed forward back propagation neural network, which is a very popular model in biological and biomedical applications, is used. This type of neural network configuration does not have feedback connections, but errors are propagated back during training using least mean squared error. The back propagation neural network is a multi-layer, feed-forward supervised learning, which requires pairs of input and target vectors. A feed forward neural network can consist of three layers, namely, (1) an input layer, (2) a number of hidden layers, and an output layer. The input layer

and the hidden layer are connected by synaptic links called weights and likewise the hidden layer and output layer also have connection weights.

The input layer contains 16 neurons representing the 16 extracted features. The output layer contains 5 neurons which represents the WBC types. It was found that 10 nodes in a single hidden layer are adequate to reach a minimum error (less than 10^{-4}). The learning rate is 0.35 and number of epochs is set to 1000.

II. THE EXPERIMENTAL RESULTS

To illustrate our proposed procedure, 40 microscopic images are shot from a stained blood smear. The WBCs are then segmented from those images, where the total sub-images that indicate the whole WBC types are found to be 100 sub-images.

In order to classify the WBCs into five classes, the target value for the WBC neural network classifier is considered as shown in Table 1. Firstly, 50 sub-images are selected for network training, while the other 50 sub-images are used for testing.

The classification test results are illustrated in Table 2. From this table, it is clear that the overall correct classification is 96%, while 4% being the overall false classification. The percentages of the correct classification for each WBC type are shown in Table 3.

Finally, Table 4 illustrates the classification performance of our proposed technique compared with some similar researcher work published in literature. In this table, it is evident that our technique gives better overall correct classification for the considered five types of white blood cells.

TABLE I. NEURAL NETWORK TARGET VALUES FOR WBC CLASSIFICATION

WBC Classes	WBC classifier targets				
	T ₁	T ₂	T ₃	T ₄	T ₅
Basophil	1	0	0	0	0
Eosinophil	0	1	0	0	0
Lymphocyte	0	0	1	0	0
Monocyte	0	0	0	1	0
Neutrophil	0	0	0	0	1

III. CONCLUSIONS

The MLP trained by Back Propagation (BP) algorithm have been used to classify five types of WBC, namely, (1) neutrophils, (2) basophils, (3) eosinophils, (4) lymphocytes, and (5) monocytes. The 16 features are used as an input to the neural network. These features are categorised as shape features (the seventh moment invariant), intensity features (the

mean, standard deviation, skewness, kurtosis of the intensity histogram), and textural features (the energy, the entropy, the correlation, the contrast, the homogeneity, and the mean and variance for each colour). The choice of features and the type of classifier play a significant role in classification accuracy results. With the above selected feature and the proposed neural network classifier a 100% classification accuracy have been obtained for the neutrophils, lymphocytes, and basophils types of WBC, while 90% accuracy have been obtained for

the other two types. For comparison purposes, the achieved results have been compared with other relative research work. This is clearly demonstrated by the results of Table 4. It is clear that the proposed neural network classifier has better classification accuracy with less number of features relative to the number of classified WBC types. As a future work, the authors would like to focus on using these WBC and RBC images to increase the capability of diagnosing some popular regional blood diseases.

TABLE II. THE NEURAL CLASSIFICATION RESULTS

#	Neural Network Output Results					Classified WBC subclass	Obtained Classification Ratio
	O ₁	O ₂	O ₃	O ₄	O ₅		
1	0.0011	0.9909	0.0002	0.0021	0.0013	Eosinophil	Correct Classification 96%
2	0.0001	0.9988	0.0002	0.0011	0.0093	Eosinophil	
3	0.0019	0.0041	1.0000	0.0031	0.0032	Lymphocyte	
4	0.0021	0.0054	0.9827	0.0002	0.2389	Lymphocyte	
5	0.0103	0.0048	0.9899	0.0028	0.0101	Lymphocyte	
6	0.0000	0.0012	0.0000	0.9972	0.0021	Monocyte	
7	0.0086	0.0061	0.9899	0.0000	0.0315	Lymphocyte	
8	0.0012	0.0000	0.9915	0.0027	0.0015	Lymphocyte	
9	0.0000	0.0073	0.0000	0.0000	0.9944	Neutrophil	
10	0.0005	0.0049	0.0012	0.0009	0.9987	Neutrophil	
11	0.0032	0.0023	1.0000	0.0000	0.0018	Lymphocyte	
12	0.0045	0.0011	0.9897	0.0001	0.0096	Lymphocyte	
13	0.1051	0.0012	0.0000	0.9919	0.0025	Monocyte	
14	0.0058	0.0093	0.0001	0.9815	0.0007	Monocyte	
15	0.0019	0.0011	0.0027	0.0125	0.9985	Neutrophil	
16	0.0004	0.0073	0.0008	0.0002	0.9917	Neutrophil	
17	1.0000	0.0001	0.0028	0.0027	0.0013	Basophil	
18	0.9959	0.0000	0.0002	0.0000	0.0028	Basophil	
19	0.0016	0.9794	0.0000	0.0081	0.0227	Eosinophil	
20	0.0000	0.9899	0.0002	0.0000	0.0036	Eosinophil	
21	0.0001	0.0052	0.9997	0.0241	0.0104	Lymphocyte	
22	0.0002	0.0028	0.9998	0.0001	0.0091	Lymphocyte	
23	0.0015	0.0093	0.0001	0.9815	0.0008	Monocyte	
24	0.0028	0.0207	0.0001	0.9798	0.0007	Monocyte	
25	0.0003	0.0004	0.0001	0.0001	0.9933	Neutrophil	
26	0.0015	0.0161	0.0014	0.1453	0.8199	Neutrophil	
27	1.0000	0.0000	0.0027	0.0001	0.0136	Basophil	
28	0.0014	0.0015	0.9977	0.0012	0.0035	Lymphocyte	
29	0.0001	0.9798	0.0021	0.0001	0.0234	Eosinophil	
30	0.0000	0.0632	0.0000	0.0000	0.8102	Neutrophil	
31	0.0003	0.0052	0.9997	0.0241	0.0051	Lymphocyte	
32	0.0026	0.0005	0.9882	0.0000	0.0225	Lymphocyte	
33	0.0003	0.0046	0.0000	0.9860	0.0815	Monocyte	
34	0.0100	0.0008	0.0000	0.9975	0.0009	Monocyte	
35	0.0017	0.0168	0.0016	0.0002	0.9579	Neutrophil	
36	0.1078	0.0092	0.0000	0.0000	0.9851	Neutrophil	
37	1.0000	0.0109	0.0043	0.0001	0.0003	Basophil	
38	0.9864	0.0075	0.0000	0.0002	0.0081	Basophil	
39	0.0107	0.9952	0.0000	0.0021	0.0006	Eosinophil	
40	0.0000	0.9979	0.0038	0.0007	0.0075	Eosinophil	
41	0.0000	0.0216	1.0000	0.0001	0.0064	Lymphocyte	
42	0.0043	0.0061	0.0000	0.9176	0.0012	Monocyte	
43	0.0011	0.0129	0.0057	0.9935	0.0068	Monocyte	
44	0.0019	0.0025	0.0000	0.9954	0.0131	Monocyte	
45	0.0008	0.0004	0.0026	0.0000	0.9974	Neutrophil	
46	0.0004	0.0012	0.0000	0.0000	0.9857	Neutrophil	
47	0.0002	0.0001	0.0000	0.0107	0.9992	Neutrophil	
48	0.0041	1.0000	0.0002	0.0108	0.0003	Eosinophil	
49	0.7911	0.0154	0.1389	0.0106	0.5477	Monocyte	
50	0.0212	0.6218	0.1000	0.2568	0.0013	Eosinophil	

TABLE III. THE PERCENTAGE OF CORRECT CLASSIFICATION FOR EACH WBC TYPE

Reference	Classification Procedure	Classification Objective	Features	Classification Accuracy
Proposed techniques	MLP-BP	Five types of WBC	16	96%
Ref. [14]	MLP-CGD	twelve types of WBC	69	89.74%
Ref. [16]	MLP-BP	lymphoblast, myeloblast and normal cell	42	93.82%

TABLE IV. CLASSIFICATION PERFORMANCE COMPARISON

CLASS NAME	NO. OF SUB-IMAGES	NO. OF MISCLASSIFIED	ACCURACY (100%)
NEUTROPHIL	10	0	100
EOSINOPHIL	10	1	90
LYMPHOCYTE	10	0	100
BASOPHIL	10	0	100
MONOCYTE	10	1	90
OVERALL PERCENTAGE OF CORRECT CLASSIFICATION			96

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", *GSTF-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., Leblebicioğlu 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.

[18] M. Darshana. and Asim, B.," Discrete Wavelet Transform Using Matlab", *International Journal of Computer Engineering and Technology*, Vol. 4, Issue 2, 2013.