An Improved Computer Vision Method for White Blood Cells Detection

Chetana K. Rane[1], Prof. Dinesh D. Patil[2]

[1] *M.E. Student, Dept. of Computer Science and Engineering, SSGBCOET, Bhusawal, Maharashtra, India*
[2]*HOD, Computer Science and Engineering, SSGBCOET, Bhusawal, Maharashtra, India*

ABSTRACT

The analysis of blood cells in microscope images can provide useful information concerning the health of patients. There are three major blood cell types, namely, erythrocytes (red), leukocytes (white), and platelets. Manual classification is time consuming and susceptible to error due to the different morphological features of the cells. This paper presents an intelligent system that simulates a human visual inspection and classification of the three blood cell types. The proposed system comprises two phases: The image preprocessing phase where blood cell features are extracted via global pattern averaging, and the neural network arbitration phase where training is the first and then classification is carried out. Experimental results suggest that the proposed method performs well in identifying blood cell types regardless of their irregular shapes, sizes and orientation, thus providing a fast, simple and efficient rotational and scale invariant blood cell identification system which can be used in automating laboratory reporting.

Keyword: *Blood cells, Intelligent identification, Neural network, Pattern averaging*

1. INTRODUCTION

There are three major cellular constituents of the blood: first, Erythrocytes or red cells which are non-nucleated biconcave diskettes with a diameter of about 8 μm. The red cells make up about 48% of the blood volume, and carry oxygen and carbon dioxide around our body. Second, Leukocytes or white blood cells which are nucleated cells with diameters ranging from 6 to 20 μm. Normal blood contains 4000–10000 leukocytes/μl of blood. The white cells play a vital role in the immune system; where they eliminate germs such as bacteria and viruses, and fight cancer cells and other toxic substances. Thirdly, Platelets which are cytoplasmic fragments of large cells called megakarocytes. They have a diameter of about 2–4 μm and normal blood containing 150,000–350,000 platelets/μl of blood. The most important function of the platelets is thrombosis and control of bleeding [1].

The three blood cell types can be differentiated from each other by their different sizes and different morphological features, such as the presence or absence of a nucleus in the cells and the shape of the nucleus. Additionally, other differentiating features can be used, such as area, eccentricity, compactness, area of central pallor (for red cells), nucleus position, number of nuclear lobes, nucleus cytoplasm ratio and color of nucleus and cytoplasm [2].

The process of automatic blood cell classification involves acquisition, detection, feature extraction, and classification. During acquisition, the blood smear is magnified to a suitable scale under the microscope, and then transformed into a digital image using a modern charge-coupled device (CCD) camera. In detection, cell segmentation yields a number of single-cell images, and each single-cell image is segmented into three regions: cell nucleus, cytoplasm, and background. During feature extraction each segmented cell is analyzed to form a feature vector from color, shape, and texture features. Finally, in classification, each blood cell is labeled by the classifier according to its feature vectors [3].

Recent studies have suggested different methods for blood cell image segmentation, which is part of the cell detection process. Buxton and Abdallahi [4] proposed a multidimensional extension of the Otsu algorithm for identifying red blood cells. Scotti [5] suggested a method for enhancing blood cell microscope images by removing the undesired microscope background, and suggested a segmentation strategy to identify white cells. Wang et al. [6] proposed a grayscale image segmentation algorithm, which combines edge-based and region-based techniques

through the morphological algorithm of watersheds, and applied their method to recognize and classify different categories of normal blood cells. Fang et al. [7] used a neural network for white blood cell image segmentation, where they trained their network using particle swarm optimization prior to recognition and classification. However, detection process and segmentation are not considered. They only focused on the feature extraction and classification processes, and used single-cell images for the development and implementation.

Recent research on feature extraction and classification of blood cells aimed at identifying the three major cell types or the identification and counting of white cells. Theera-Umpon and Dhompongsa [4] proposed a method for the classification of white blood cells using only their nucleus information. Using single-cell images, they analyzed a set of white blood cell nucleus-based features using mathematical morphology, and classified the white cells using Bayes classifiers and a neural network. Mircic and Jorgovanovic [6] focused their work also on classifying white cells, and used binary single-cell images and a neural network classifier for their purpose. Shitong and Min [3] suggested an automatic white blood cell detection system, where binary threshold segmentation is applied to the cell images and classification is performed using a fuzzy cellular neural network. However, despite the successful results that were demonstrated by these recent methods, they only targeted the identification of white blood cells.

The classification of the three blood cell types using neural networks was also addressed in previous studies. Sheikh et al. [1] suggested using wavelet transforms for feature extraction of single-cell images, and a neural network classifier. Lin et al. [2] used an edge detection method and a neural network for classifying the three types of blood cells. Markiewicz et al. [1] considered the features of images of the blood cells related to the texture, geometry and histograms, and used a support vector machine as the classifier. Zheng et al. [6] proposed the use of cell image data as direct input into neural networks to determine the feasibility of direct classification by using pixel intensity information, and they classified images using three-layer and four-layer neural networks based on the back propagation learning algorithm.

However, all these methods are computationally expensive and more optimized for identifying certain cell types. Additionally, none of these methods have considered for modeling the way human experts identify blood cell types from microscopic images.

This paper presents a fast, simple and efficient Intelligent Blood Cell Identification System (IBCIS) that simulates a human expert recognition of the three blood cell types using global pattern averaging and a neural network. The human expert would normally "look" at the blood cell image and then quickly "identify" the type, regardless of the cell's different morphological features or the size and rotation. Our hypothesis is that the "look at a cell" can be approximated in machines using global pattern averaging, whereas, the "identify the type" can be simulated by a trained neural network. The proposed system will be implemented using 360 single-cell images; comprising 120 images of each of erythrocytes (red), leukocytes (white) and platelets blood cells, in various shapes, sizes and rotations. Successful results have been achieved.

2. BLOOD CELL IMAGE DATABASE

In general, databases that are used in developing blood cell classification systems rely on microscopic cell images. The variety of information in these images makes cell identification difficult for machines due to the different sizes, shapes and colors. However, as our novel approach to cell identification is based on simulating a human expert's visual recognition, who is normally able to identify the three types of cells despite the above problems, the different morphological features, sizes and rotations of cells are not considered as an obstacle in this work and are left to the neural network to learn via global pattern averaging.

The development and implementation of the proposed intelligent identification system uses 360 single-cell images representing 90 different blood cells (30 red, 30 white, and 30 platelets). Fig. 1 shows examples of the different blood cells. Each cell has been rotated by 90° and its image is stored thus producing four images for each cell (at 0°, 90°, 180° and 270°), as shown in Fig. 2. The rotations are aimed at testing the trained system's rotational invariance capability. The obtained original color images were then converted from RGB to gray level, and resized to 70 × 70 pixels, in preparation for the feature extraction phase.

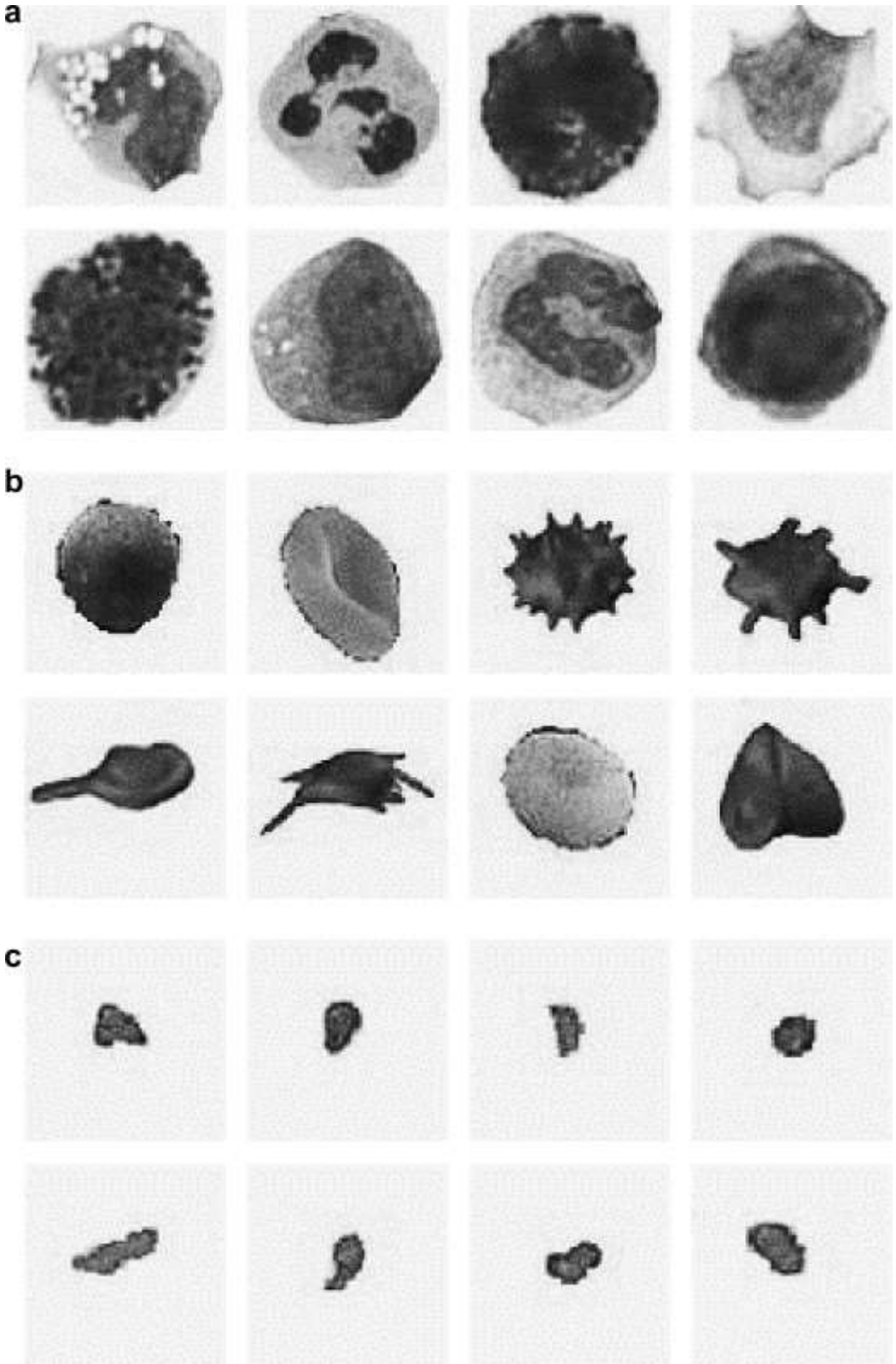

Fig -1: Blood cell examples of varying shapes and sizes. (a) Red blood cells; (b) white blood cells; (c) platelets blood cells.

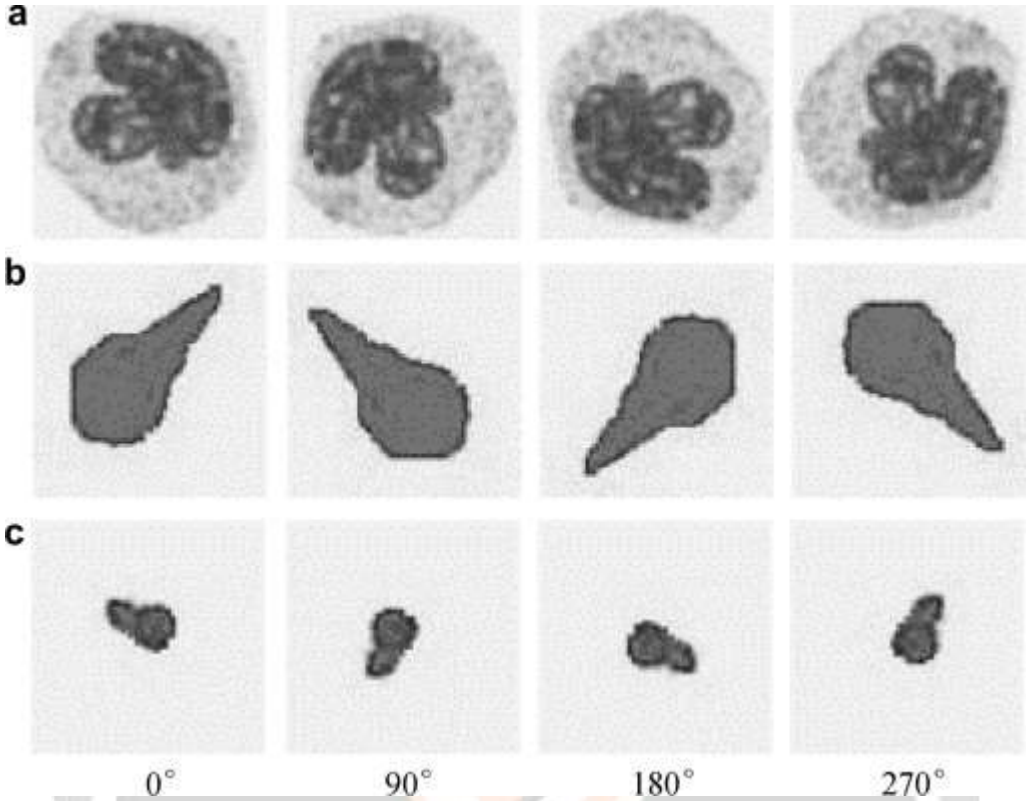

Fig -2: Examples of blood cell rotations by 90°. (a) Red; (b) white; (c) platelets.

3. THE IDENTIFICATION SYSTEM

IBCIS consists of two phases: a feature extraction phase where single-cell images undergo global pattern averaging in preparation to be presented to the second phase, which is identified by a neural network. Once the neural network converges and learns, the second phase consists only of one forward pass that yields the identification result.

3.1 Feature Extraction Phase

This phase is a data preparation phase for neural network training and classification. The extracted features from the blood cell images will be used as the input to the neural network. Global pattern averaging is used to extract the feature vectors from the cell images, which in our hypothesis approximates the human expert's visual inspection of the blood cell.

A single-cell image, which is gray and of size (70 × 70) pixels, is segmented and the values of the pixels within each segment are averaged. The resulting average values are then used as input data for the neural network. The averaging of the segments within an image reduces the amount of data required for neural network implementation thus providing a faster recognition system. Global pattern averaging can be defined as follows:

equation(1)

$$PatAv_i = \frac{1}{s_k s_l} \sum_{l=1}^{s_l} \sum_{k=1}^{s_k} p_i(k,l)$$

\

where k and l are segment coordinates in the x and y directions, respectively; i is the segment number; Sk and Sl are segment width and height, respectively; Pi(k,l) is pixel value at coordinates k and l in segment i; PatAvi is the average value of pattern in segment I , which is presented to neural network input layer neuron i. The number of

segments in each image (of size XY pixels; X = Y = 70) contains a cell, and the number of neurons in the input layer is i, where
equation(2)
I={0,1,2,…,n}

and
equation(3)
n=(X/sk)(Y/sl)

Segment size of 5 × 5 pixels (Sk = Sl = 5) was used and average values representing the image were obtained, thus resulting in 196 average values in total (n = 14) that were used as the input to the neural network for both training and testing. Previous work using this pre-processing method showed that sufficient representation of the objects within the images and meaningful data within the averaged patterns were obtained to aid the neural network learning and classification [7]. Pattern averaging provides meaningful learning and marginally reduces the processing time. For the work presented within this paper, global pattern averaging overcomes the problem of varying pixel values within the segments as a result of rotation, thus, providing a rotation invariant system. A segment size of 5 × 5 pixels is used, resulting in a 14 × 14 bitmap of averaged pixel values that will be used as the input for the second phase, namely neural network training and generalization.

3.2 Classification Phase

During this phase a supervised neural network is used. The neural network is based on the back propagation learning algorithm due to its implementation simplicity, and the availability of sufficient database for training this supervised learner. The neural network consists of an input layer with 196 neurons, one hidden layer with 40 neurons and an output layer with 3 neurons.

Training the neural network uses 60 non-rotated blood cell images (20 red, 20 white and 20 platelets). The remaining 300 blood cell images are not exposed to the neural network during training, and will be used to generalize or test the trained network. The hidden layer of the neural network contains 40 neurons which assures meaningful training while keeping the time cost to a minimum. The output layer has three neurons corresponding to the three blood cell types. Binary output data representation is used as follows: red (1 0 0), white (0 1 0), and platelets (0 0 1).

During the learning phase, the number of hidden layer neurons, the learning coefficient, and the momentum rate were adjusted during various experiments in order to achieve the required minimum error value of 0.003 which was considered to be sufficient for this application. Fig. 3 shows the topology of this neural network within the intelligent system.

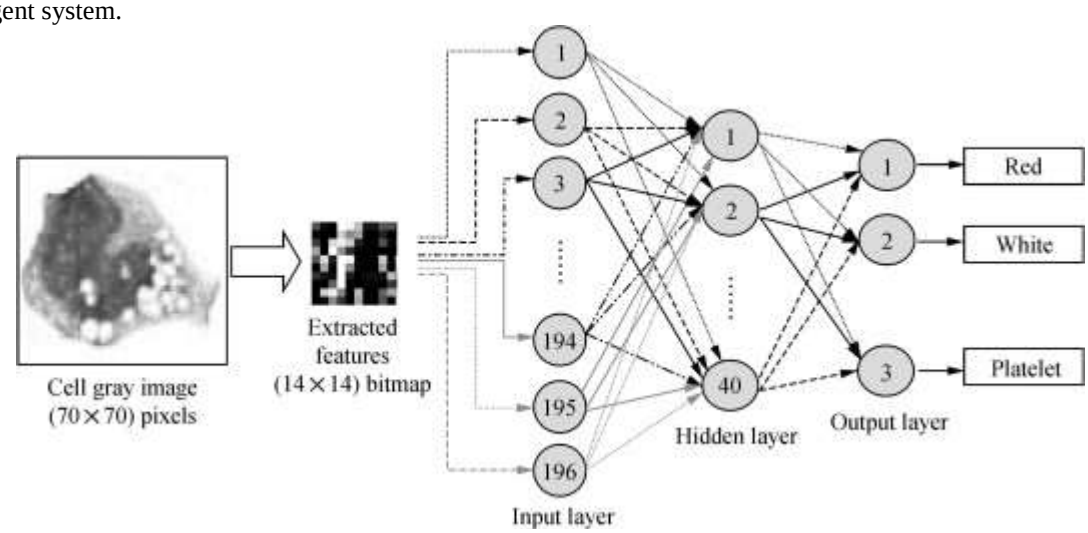

Fig. 3: Global pattern averaging and neural network topology.

4. IMPLEMENTATION RESULTS

The neural network learnt and converged after 3707 iterations and within 119.6 s, whereas the running time for the generalized neural network after training and using one forward pass was 0.016 s. These results were obtained using a 2.8 GHz PC with 512 MB of RAM, Windows XP OS and Borland C++ compiler. Table 1 lists the final parameters of the successfully trained neural network, whereas Fig. 4 shows the error versus iterations graph during training. The robustness, flexibility and speed of this novel blood cell identification system have been demonstrated through this application.

Parameter	Final value
Input layer nodes	196
Hidden layer nodes	40
Output layer nodes	3
Learning coefficient	0.015
Momentum rate	0.31
Random initial weights	(−0.45) to (+0.45)
Minimum error	0.003
Iterations	3707
Training time (s)	119.6
Run time (s)	0.016

Table 1: Neural network final training parameters

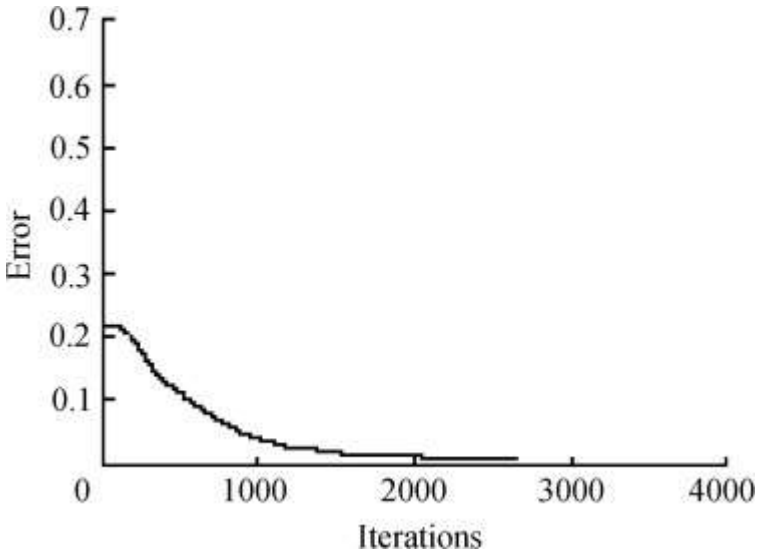

Fig. 4: Error vs. iterations training curve.

The implementation results of the trained intelligent system were as follows: using the training image set (60 non-rotated blood cell images) yielded 100% recognition as expected. The intelligent system implementation using the testing image set (300 images that were not previously exposed to the neural network) yielded correct blood cell identification of 297 images, thus achieving an impressive 99% correct identification rate. Combining results using testing images (300) and training images (60), yielded an overall correct identification rate of 99.17%. Table 2 shows the intelligent blood cell identification results in details.

Blood cell type	Image set	Correct identification rate
Red	Training	20/20 (100%)
	Testing	99/100 (99%)
White	Training	20/20 (100%)
	Testing	98/100 (98%)
Platelet	Training	20/20 (100%)
	Testing	100/100 (100%)
All blood cells	Training	60/60 (100%)
	Testing	297/300 (99%)
	Total	357/360 (99.17%)

Table 2: Intelligent blood cell identification results

There were only three incorrect identifications; these were as follows: one red cell identified as white, one white cell identified as red, and another white cell identified as platelet. Fig. 5 shows these incorrect identifications.

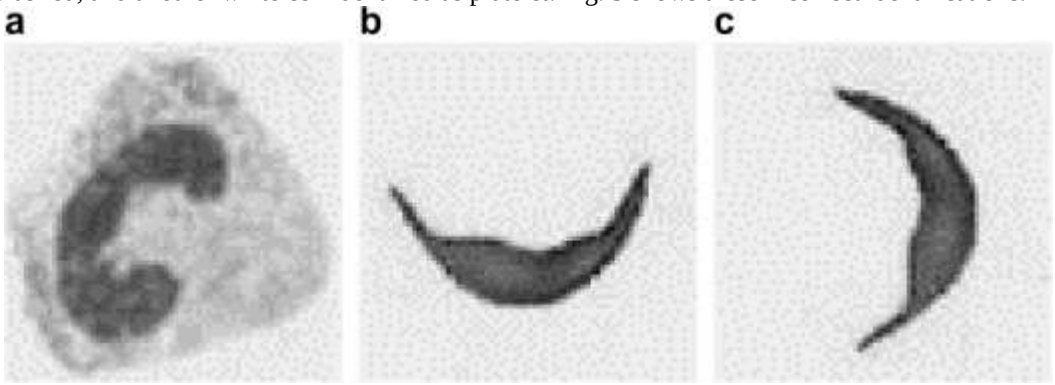

Fig. 5: Incorrect identifications. (a) A red cell identified as white; (b) a white cell identified as platelet; (c) a white cell identified as red.

5. CONCLUSION

A novel approach to blood cell identification, based on simulating a human expert's "look" and "identify", has been introduced in this paper. The "look" effect is approximated via global pattern averaging. When we, humans, have a quick "look" at a familiar object (blood cell types are familiar to a human expert), we do not observe the detailed features but rather a general global impression of the object. This in our hypothesis can be applied to identifying blood cells which usually have irregular shapes, different sizes and colors. The "identify" effect is simulated by training a simple but efficient neural network. The ability of the trained neural network to identify blood cells, despite the irregular non-uniform shapes, is due to training the network using feature approximations or "fuzzy" feature vectors rather than using "crisp" feature vectors. The averaged patterns are true representations of a cell image regardless of its size or orientation.

The system presented in this paper was implemented using 360 single-cell images of the three major blood types (red, white and platelets). The 360 images represented 90 different blood cells; each rotated by 90°, thus providing four different orientation for each cell. The images of the rotated cells were only used for testing the trained neural network, and to demonstrate the proposed system's rotational invariance.

An overall correct identification rate of 99.17% was obtained. The robustness and success of this simple but efficient blood cell identification system was further demonstrated by its quick run time (one neural network forward pass) of 0.016 s. Time cost was kept minimal through image preprocessing and reduction of input/hidden

layer neurons in the topology of neural network. The quick run time was achieved using a 2.8 GHz PC with 512 MB of RAM, Windows XP OS and C++ programming language. This time cost can be further improved by using a faster system.

Future work includes enlarging the blood cell database, and incorporating an intelligent white blood cell counter within the identification system.

6. REFERENCES

[1] H. Sheikh, B. Zhu, E. Micheli-Tzanakou, Neural networks and blood cell identification

[2] Lin W, Xiao J, Micheli-Tzanakou E., A computational intelligence system for cell classification. In: IEEE international conference on information technology applications to biomedicine, May 1998, p. 105–9.

[3] S.T. Wang, M. Wang, A new detection algorithm (NDA) based on fuzzy cellular neural networks for white blood cell detection
IEEE Trans Inform Technol Biomed, 10 (1) (2006),

[4] Buxton BF, Abdallahi H, Fernandez-Reyes D, et al. Development of an extension of the Otsu algorithm for multidimensional image segmentation of thin-film blood slides. In: IEEE international conference on computing: theory and applications, March 2007,

[5] Scotti F. Robust segmentation and measurements techniques of white cells in blood microscope images. In: IEEE instrumentation and measurement technology conference, April 2006

[6] Wang W, Song H, Zhao Q. A modified watersheds image segmentation algorithm for blood cell. In: IEEE international conference on communications, circuits and systems, vol. 1. 2006,

[7] Fang Y, Zheng CX, Pan C, et al. White blood cell image segmentation using on-line trained neural network. In: IEEE conference on engineering in medicine and biology, Sept 2005