Sickle Anemia and Distorted Blood Cells Detection Using Hough Transform Based on Neural Network and Decision Tree

Hany A. Elsalamony[1]
[1]Mathematics Department, Faculty of Science, Helwan University, Cairo, Egypt.
h_salamony@yahoo.com

Abstract - *Sickle-cell anemia is one of the most important common types of anemia disease. This paper presents proposed algorithm in two parts, one is the construct an algorithm can detecting and counting RBCs (benign or distorted) in a microscopic colored image; even if they are hidden or overlapped. Second part is checking and analysing the constructed data of RBCs by applying the most common important techniques in data mining; neural network and decision tree. The experimental results are demonstrated high accuracy, and success using these two models in predicting for all the benign or distorted cells. This algorithm has achieved the highest segmentation by about 99.98% of all input cells, which is contributed to improve the diagnosis of Sickle anemia. The Neural Network has agreed with the detection by the proposed algorithm in prediction outcome by about 96.9%, whereas the Classification and Regression (C&R) tree has achieved 92.9%.*

Keywords: Sickle Anemia; Image watershed segmentation; Red Blood Cells' detection and counting; C&R tree; Neural Network.

1 Introduction

Nowadays, the analysis of the blood cells' microscopic image is very impressive diagnostic tool for many diseases. Actually, human blood is a complex combination of plasma, red blood cells (RBCs), white blood cells (WBCs), and platelets. Plasma is the fluid component, which is contained melted salts and proteins. RBCs make up about 40% of blood volume. WBCs are less, but greater in size than RBCs. The platelet cells are similar particles, which are smaller than WBCs and RBCs [16].

Anemia is occurred when the blood has a lower than the normal number of red blood cells (RBCs) or they have not been enough hemoglobin. RBCs have been made inside the larger bones of the body in the spongy marrow. Bone marrow is always making new red blood cells to replace old ones. Normal RBCs die after they living 120 days in the bloodstream. Their job is carrying oxygen and removing carbon dioxide (a waste product) from the body. In addition, RBCs are disc-shaped and moving easily through blood vessels. They contain an iron-rich protein called hemoglobin. This protein transmits oxygen from the lungs to the rest of the body [16]. In Sickle-cell anemia, the body makes sickle-shaped red blood cells in a serious disorder. Sickle cells contain abnormal hemoglobin called Sickle hemoglobin or hemoglobin S, which it helps the cells to develop a sickle, or crescent, shape. In fact, Sickle cells are very dangerous because rigidity and sticky of them, that are tended to block blood flow in the blood vessels of the limbs and organs. The blocked blood flow can cause pain, organ damage, and can raise the probability of infection. Moreover, the abnormal Sickle cells usually die after only about 10 to 20 days. The bone marrow cannot make new red blood cells fast enough to replace the dying ones [16]. Figure.1 illustrates the danger of Sickle anemia and kinds of it.

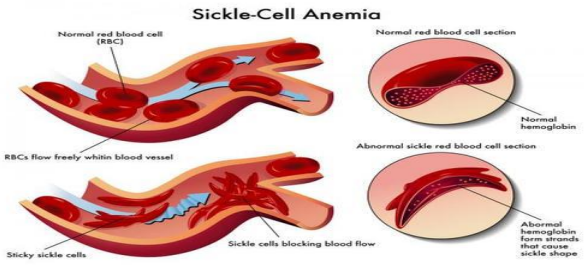


Figure.1 Kinds of Sickle's anemia and RBCs

In addition, the Sickle-cell anemia is most common in people whose families derived from Mediterranean countries, Africa, south or Central America, especially Panama, Caribbean islands, Saudi Arabia, and India. The United States estimated that the infection from 70,000 to 100,000 people, mainly African Americans. The main discovery of this disease is depending on blood test analysis that can detect Sickle cells.

Recently, the microscopic image analysis helps as an impressive diagnostic tool for the infected blood cells' detection [16]. The Hough transform is the most important technique in image analysis and segmentation to help in that way. It is depending on extracting features related through the segmentation process of microscopic image. Hough transform that is generally used today was invented by Richard Duda and Peter Hart in 1972, who called it a "generalized Hough transform" after the related 1962 patent of Paul Hough [8], [12].

On the other hand, Data mining techniques have gained popularity and useful for illustrative and predictive applications of image analysis. Two techniques are applied to the resulted data on blood cells' detection. Neural network (NN) is one of these techniques; it has been applied successfully in the identification and control of dynamic systems. Previously, the progress of for NN began by David Rumelhart in 1986. He presented the back propagation that distributed pattern recognition errors and using many layers throughout the network.

Another one of data mining applied techniques is the classification and regression tree (C&R) as the most common and powerful technique for the classification and prediction in the decision tree (DT). It can generate understandable rules, and to handle both continuous and categorical variables [7]. Historically, the seminal book by Bremen et al. (1993) provided an introduction to decision trees that is still considered the standard resource on the topic. The C&R tree algorithm popularized by Breiman, Friedman, Olshen, and Stone in 1984, and by Ripley in 1996 [3]. Two reasons behind the popularity of decision tree techniques are (1) the procedures are relatively straightforward to the understand and explain, and (2) the procedures address a number of data complexities, such as nonlinearly and interactions, that commonly occur in real data [7].

The work of this paper is divided into two parts; one is applying a proposed algorithm to detect the benign and distorted blood cells (Sickle anemia) and counting them depending on segmentation of their shapes using Circular Hough Transform, watershed, and morphological tools. The second part is introducing an algorithm for checking and analysis of the resulted cells' data variables (Areas, Convex Area, Perimeter, Eccentricity) by applying the common most important techniques in data mining; neural network and decision tree (classification and regression tree) to get the right decision for diagnoses [2].

The rest of the paper is organized as follows; Section (2) focuses on the related work, and the definition and features of Hough transform is presented in section (3). Section (4) overview of the neural network. Section (5) is presented an overview of classification and regression of decision trees. The proposed algorithm is discussed in section (6). The experimental results show the effectiveness of each model in section (7). The conclusion has been presented in section (8).

2 Related work

In recent years, research on blood cells' recognition and diagnosis of diseases has grown rapidly. In May 2013, K. Thirusittampalam, M. J. Hossain, O. Ghita, and P. F. Whelan, developed a novel-tracking algorithm that can extract the cell motility indicators and determined the cellular division (mitosis) events in large time-lapse phase-contrast image sequences. Their process of automatic unsupervised cell tracking carried out in a sequential manner, where the inter frame cell's association is achieved by assessing the variation in the local cellular structures in consecutive frames from the image sequence. The experimental results indicated that their algorithm achieved 86.10% overall tracking accuracy and 90.12% mitosis detection accuracy [9].

Another proposal introduced by H. A. Khan and G. M. Maruf in May 2013, that presented an algorithm for cell segmentation and counting by detection of cells' centroids in microscopic images. The method is specifically designed for counting circular cells with a high probability of occlusion. The experimental results showed an accuracy of 92% of cell counting even at a very high 60% overlap probability [5].

An algorithm presented by M. C. Mushabe, R. Dendere and T. S. Douglas in July 2013, that they were identified and counted red blood cells (RBCs) as well as parasites in order to perform a parasitemia calculation. Morphological operations and histogram-based threshold were used to detect the red cells. They used boundary curvature calculations and Delaunay triangulation to split overlapped red cells. The parasites are classified by Bayesian classifier with their RGB pixel values as features. The results showed 98.5% sensitivity and 97.2% specificity for detecting infected red blood cells [10]. The next section introduces an overview on Hough transform.

3 Hough Transform

The Hough Transform is a popular feature extraction technique that converts an image from its Cartesian to its Polar coordinates. Any point within the image space is represented by a sinusoidal curve in the Hough space. In addition, two points to a line segment generate two curves, which are overlaid at a location corresponds with a line through the image space. Even though this model form is very easy it is deeply complicated for the case of complex shapes due to noise and shape imperfection, also the problem of finding slopes of vertical lines. Circular Hough Transforms (CHT) solved this problem by putting a transformation of the centroid of the shape in the x-y plane to the parameter space [6].

However, there are three essential steps, which are common to all CHT: first one is an Accumulator Array Computation, which is working as that foreground pixel of a high gradient are chosen as being candidate pixels and are allowed to 'votes' in the accumulator array. Center estimation is the second step; the circle centers are expected by detecting the peaks in the accumulator array by voting of candidate pixels that are belonging to an image circle tend to accumulate in the accumulator array box corresponding to the circle's center.

Figure.2 shows an example of the candidate pixels (solid dots) falling on an actual circle (solid circle), and their voting patterns (dashed circles) which coincide with the center of the substantial circle. The third step in CHT is radius estimation; that is if the same accumulator array has used for more than one radius value, as is commonly done in CHT algorithms, the radii of detected circles have estimated as a separate step [8].

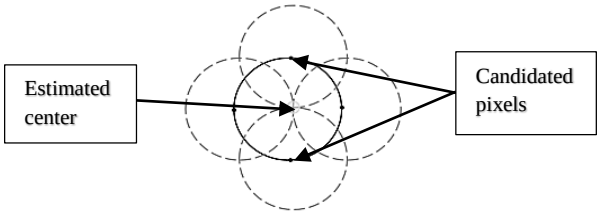


Figure.2 Circle center estimated.

The radius has estimated clearly by using radial histograms; however in Phase-Coding, the radius can be estimated by simply decoding phase information from the estimated center located within accumulator array [8]. This paper has used CHT to detect and count RBCs; even if they are hidden or overlapped. Watershed and morphological functions are used for enhancing and separating overlapped cells during the segmentation process.

4 Overview of Neural Network

At the present time, Neural Network (NN) with back propagation is the most popular artificial neural network construction. It is known as a powerful function approximation for prediction and classification problems. Historically, NN is a mutually dependent group of artificial neurons that uses a mathematical model for information processing with a connected approach to computation by Freeman in 1991 [4], [7]. The NN structure is organized into layers of input, output neurons, and hidden layers. The activation function may range from a simple threshold function, or a sigmoid, hyperbolic tangent, or radial basis function [1].

$$y_i = f\left(\sum w_{ij} x_i\right) \quad (1)$$

The back propagation is a common training technique for NN. This training process is to perform a particular function by adjusting the values of the connections (weights) between elements [2], [17]. Actually, three important issues in NN need to be addressed; selection of data samples for network training, selection of an appropriate and efficient training algorithm, and determination of network size [14], [15]. Moreover, NN has many advantages, such as the good learning ability, less memory demand, suitable generalization, fast real-time operating, simple and convenient to utilize, suited to analyze complex patterns, and so on. On the other hand, there are some disadvantages like that it requires high-quality data; variables must be carefully selected a priori, the risk of over-fitting, and requires a definition of architecture [4].

5 Overview of Classification and Regression Tree

Classification and regression trees (C&R) are the most common and popular non-parametric decision tree learning technique. In this paper, a regression tree only uses for numeric data values. C&R builds a binary tree by splitting the records at each node according to a function of a single input variable. The measure that used to evaluate a potential splitter is diversity. The best splitter is the one that decreases the diversity of the record sets by the great one. This method uses recursive partitioning to split the training records into segments with similar output variable values [7]. Moreover, the impurity that used in each node defined in the tree by two measures; entropy, as in Equation (2), and Gini (chosen in this paper).

$$Entropy(t) = -\sum_j p(j|t)\log p(j|t) \quad (2)$$

The Gini index generalizes the variance impurity, that variance is of distribution related to the two classes. However, the Gini index, as in Equation (3), can also be useful as the expected error rate if the class label is randomly chosen from the class distribution at the node. This impurity measure has been slightly stronger at equal probabilities (for two classes) than the entropy measure. Gini holds some advantage for an optimization of the impurity metric at the nodes [11].

$$g(t) = 1 - \sum_j p^2(j|t) \quad (3)$$

When the cases in a node are evenly distributed across the categories; the Gini index takes its maximum value of *1-(1/k)*, where *k* is the number of categories for the target attribute. Furthermore, for all cases in the node, which are belonged to the same category the *Gini* index equals zero. The proposed algorithm is displaying in the next section.

6 The proposed algorithm

The goal of this paper is to detect benign and distorted blood cells in a colored microscopic image and distinguishing between them. This work can help doctors, physicians, chemists… who cares about blood cell detection, analysis, and determination of diseases. The proposed algorithm is divided into two steps, as shown in Figure.3. One is applying CHT with morphological functions on bright and dark of intensity cells to detect and count benign and distorted blood cells. The second step; NN and C&R tree are applied to test and check the performance of the proposed algorithm for diagnosing and deciding that the patient has Sickle anemia or not and which it is more effective than another. The classification and prediction are depended on the detected cells' data variables (Areas, Convex Area, Perimeter, Eccentricity) to reduce the errors in detection operation and ensure about final diagnosis for Sickle Anemia existence. In the first part of CHT many operations have been carried out through it:

- Cell polarity; indicates whether the circulating blood cells are brighter or darker than the background.
- Computation method (Two-stage) is used to calculate the accumulator array of CHT. It is based on computing radial histograms; radii are clearly applying the estimated cell centers along with the image information [13].
- Sensitivity factor is the sensor of the accumulator array in CHT. The detection is including weak and partially hidden or overlapped cells; however higher values of the sensitivity increasing the risk of false detection.

- Edge gradient threshold; cells have generally a darker interior (nuclei) and surrounded by an outlying bright halo. The edge gradient threshold is very useful for determining edge pixels in these cases of image, both weak and strong blood cells based on their contrast are detected well by setting a lower value in the threshold. It detects fewer cells of weak edges by increasing the value of the threshold [1].

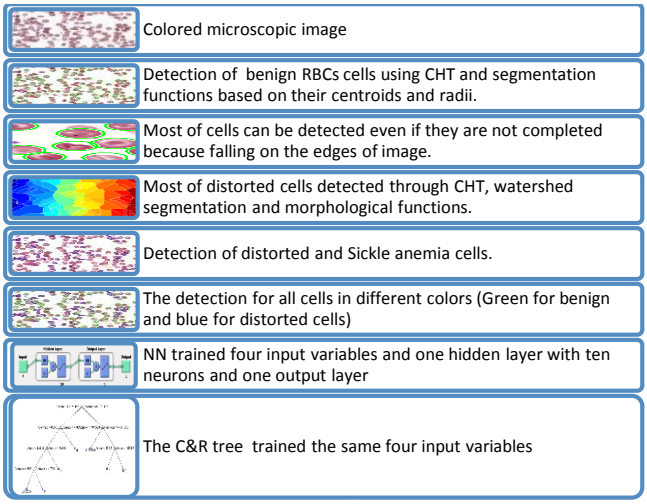


Figure.3 the main points of the proposed algorithm

The final stage is the segmentation process; is displaying of input image with all the contoured benign (green) and distorted (blue) RBCs. Through this operation; the area, convex area, perimeter, eccentricities for each cell are measured, also the good and distorted cells can be counted and diagnosed with Sickle anemia. These measures are used as input variables to train NN and C&R tree while the output (target) is measured based on the solidity (S), which is a division cell area over its convex area for all cells, as in the Equation (4):

$$S = \frac{area_c}{(convex\ area_c)} \tag{4}$$

$$T_c = \begin{cases} 1 & if \quad 0.95 < S_c \leq 1 \\ 0 & if \quad 0 \leq S_c \leq 0.95 \end{cases} \tag{5}$$

where $area_c$ is the area in each cell, and $convex\ area_c$ represents its convex area for all detected cells (benign and distorted). The target T_c have been two values 1 and 0 based on the solution of Equation (4); such that if any of solidity values S_c is greater than 0.95 up to 1 (perfect cell) T_c takes the value 1 with the decision benign. On the other hand, if T_c takes the value 0, then S_c value has less than or equal 0.95 according to Equation (4) with the decision that the cell is distorted and may be sickled.

The back propagation neural network has been trained and tested four variables as an input layer and ten neurons in the hidden layer whereas one neuron in the output layer. Moreover, three kinds of samples are applied training, validation, and testing samples. The training samples are presented in the network during the training process by 80% from all samples in the input variables, and the network is modified according to its error. Accordingly, only 10% are used for the validation samples to measure network generalization and to pause training when generalization stops improving, and the remaining 10% of all samples of cells are introduced to be a testing sample that have no effect on training and so provide an independent measure of network performance during and after training. In addition, the mean square error (MSE) is applied, that it is defined as the average squared difference between outputs and targets, whereas lower values are better, Zero means no error. Table 1 illustrates the description of the formed cells' data, which is automatically computed for all cells (benign and distorted).

TABLE 1 BLOOD CELLS' VARIABLES

#	Variables	Type	Domain
1	Areas	Ranged	163:1376
2	Convex area	Ranged	187:1781
3	Perimeter	Ranged	58.97056:218.2082
4	Eccentricity	Ranged	0:0.965444268
5	Target	Binary	(1 for benign, 0 for distorted cells)

On the other way, the recursive binary C&R tree has been applied to the term of regression because all the variables containing numeric values, as in Table 1. In this Table, all variables have ranged in different types whereas only the output (target) variable has binary values such as 1 values for benign cells and 0 values for all distorted cells. The binary tree has divided into two branches based on the Gini index recursively trained with maximum tree depth five levels and stopping when a minimum records in parent branch 2% and minimum records in child branch 1%.

Finally, the performance of each classification model is evaluated using three statistical measures: classification accuracy, sensitivity and specificity. These measures are defined as true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*). A true positive decision occurs when the positive prediction of the classifier coincided with a positive prediction of the previously segmentation. A true negative decision occurs when both the classifier and the segmentation suggests the absence of a positive prediction. False positive occurs when the system labels benign cell (positive prediction) as a malignant or distorted one. Finally, false negative occurs when the system labels a negative (malignant) cell as positive. Moreover, the classification accuracy is defined as the ratio of the number of correctly classified cells and is equal to the sum of *TP* and *TN* divided by the total number of RBCs (*N*) [9].

$$Accuracy = \frac{TP+TN}{N} \tag{6}$$

The sensitivity refers to the rate of correctly classified positive and is equal to *TP* divided by the sum of *TP* and *FN*.

$$Sensitivity = \frac{TP}{TP+FN} \tag{7}$$

Specificity refers to the rate of correctly classified negative and is equal to the ratio of *TN* to the sum of *TN* and *FP* [9].

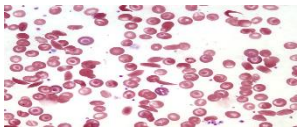
$$Specificity = \frac{TN}{TN+FP} \quad (8)$$

The next section is displaying the experimental results of the proposed algorithm in details.

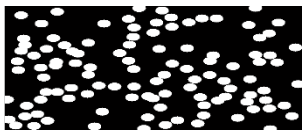
7 The experimental results

As mentioned before, the experimental results are displayed in two parts; one is concerned in the RBCs detection and segmentation to distinguish between benign and distorted cells (caused by Sickle Anemia) and counted automatically. The work of second part is the checking the previous detection process exported from part one by predicting and categorizing the segmentation results using the most famous two classification models in data mining NN and C&R tree [2]. Accordingly, the proposed algorithm has provided the ratio of errors during the detection process and performance of success using CHT in this case of RBCs detection but not effictive.

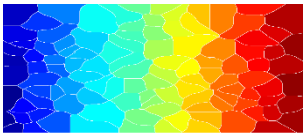
In part one; the detection process has been started by importing and reading the microscopic colored RBCs image. This detection for all cells (benign or distorted) is done using CHT, watershed process, morphological techniques for enhancing the detection process. In the same way, CHT is applied under the conditions of cell polarity to determine all dark and bright cells according to their intensity. Then, two-stage technique that used to compute the accumulator array of CHT; the sensitivity of this accumulator array of the proposed algorithm is 0.97 for brightness and 0.90 for dark cells, and Edge gradient threshold 0.2 that detect fewer cells with weak edges. Actually, these conditions help CHT to detect most of the benign RBCs (near to the circle in shape), those positioned singular, overlapped, and even those are attached to other distorted cells. Figure.4 shows the original image of RBCs in (a), and the proposed algorithm in details is illustrated from (b) to (f).



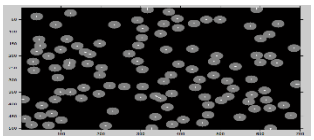
(a) Original image



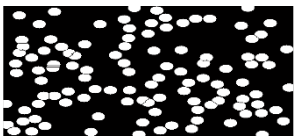
(b) Mask on only Benign cells



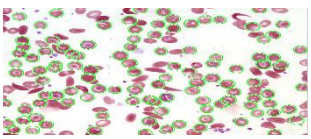
(c) Watershed on the blobs in (b)



(d) Marks on the centroids of cells



(e) Separated cell blobs

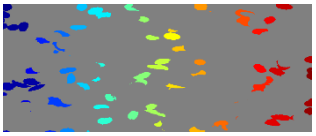


(f) Contour benign cells only

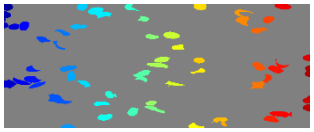
Figure.4 (a) The original image, (b) The healthy cells masked, (c) Watershed for each blob region, (d) Makes a small mark on every cell's centroid, (e) The cell's blobs separated, and (f) The final detected benign cells.

In (b); a mask of balls constructed using centroids and radii for each cell, which have determined before; in this case and after the masking process, many overlapping benign cells appeared as a one big region according to the shape and size of normal cells. For that reason, the watershed process in (c) is applied to get the optimum separation in this abnormal shape of cell. In fact, to get the optimal separation of overlapped cells their centroids are used to put marks on them as in (d). Now the separation has been ready regarding these marks for each cell blob as in (e). Finally, by applying all the previous steps; the only benign cells are contoured by the green line, extracted to count, and distinguished from the other distorted cells, as well shown in the second step of part one. Through this detection operation of benign cells; the number of these cells is equal to 109 benign cells out of 180 total number of all detected cells (benign and distorted) in this image. In fact, the remaining number of cells (71) may be considered as detection errors, distorted cells (Sickle Anemia), platelets or even WBCs. This image has not any WBCs; but if they exist in other images, the proposed algorithm can easily detect and count them. On the other hand, platelets have been neglected because the algorithm concentration only on RBCs and distorted Sickle Anemia disease.

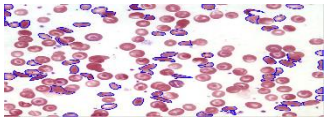
Additionally, the algorithm has been determined 177 cells as a total of detecting blood cells by 99.98% success ratio according to the image in Figure.4 (a). Therefore, the next step is trying to know that how many cells out of 71 are Sickle cells or initiated to be it. Firstly, all these 71 strange shapes (crescent, elliptic, platelets, and unknown) are discovered and displayed using the same previously steps of benign cell detection. Figure.5 illustrates last two steps of the proposed algorithm, which is applied to the distorted cells. In (a) a colored segmented image for all unknown distorted shapes with a noise like platelets … etc. The Sickle cells or the cells that initiated to be Sickle are detected without any noise in (b). On the other hand, the only deformed cells are counted 57 (Sickle or initiated to be) out of 71. The last detection of distorted cells is contoured by blue line shown in (c). In the (d) of Figure.5, the final detection and tracking of all cells (benign by green color and distorted with blue color) has been completed.



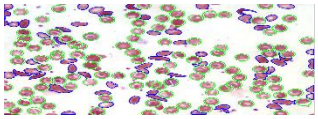
(a) Colored unknown shapes



(b) Distorted cells without noise



(c) Final detection



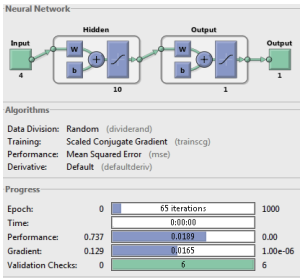
(d) Concluding detection of all cells

Figure.5 (a) All unknown shapes, (b) the distorted cells, (c) the final detection of only distorted cells, (d) the final detection of all cells.

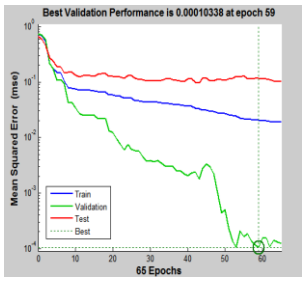
After the end of this part of detection and segmentation, the second part of the proposed algorithm is began that concerning prediction and analysis of resulted cells' detection to ensure

that the patient has exactly Sickle Anemia, and what is the ratio of effectiveness in the detection process? and what is the important variable of these infected cells that fall in the circle of concentration to give help in the diagnosing decision? These variables are the properties of cells (benign and Sickle), which are computed through the detection process. They are assigned as input variables to applying NN and C&R tree to stand upon the ratio of errors according to the presented algorithm. The input variables consist of Areas, Eccentricities, Perimeters, and Convex areas of all cells, and the target is computed as in Equation (5) depending on the solidity in Equation (4). In the previous, the back propagation has been trained 128 values as an 80% of all samples, 10% validation and testing (i.e. 16 samples for each one).

Experimentally, NN consists of four input variables, ten neurons in one hidden layer and one output layer. Actually, the network already has succeeded after 65 iterations out of 1000 as maximum iterations of the epoch, performance 0.0189, gradient 0.0165, and 6 validation checks as in Figure.6 (a). In (b) the best validation performance is shown as 0.00010338 at epoch 59, when the training in a blue line, validation in green, and the test in the red line. In the same context, the mean square error of training, validation, and testing processes are 2.059e-2, 1.03383e-2, and 1.17910e-2, respectively.



(a) The training of NN process



(b) The best validation performance

Figure.6 (a) the back propagation NN, and (b) the best validation performance.

Figure.7 shows the confusion matrices for training, validation, and testing processes. In this Figure, the predictions of NN model are compared with the original classes of the target T_c to identify the values of true positives, true negatives, false positives, and false negative. These values have been computed to construct the confusion matrix, where each cell contains the number of cases classified for the corresponding combination of desired and actual classifier outputs, and it achieved 96.9%. Accordingly, accuracy, sensitivity, and specificity are approximated the probability of the positive and negative labels to being true and have assessd the usefulness the algorithm on an NN model. Respectively, the accuracy, sensitivity, and specificity classifications of NN have achieved 98.4%, 100%, and 93.3% success of training samples.

In the same context, the validity has achieved 100%; the accuracy, sensitivity, and specificity classifications of the test samples have achieved 81.3%, 90.9%, and 60.0% respectively.

Last but not least, NN agreed with all detected cells using the proposed algorithm by about 96.9%.

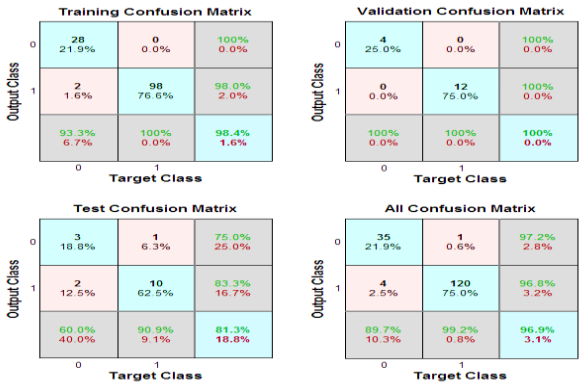


Figure.7 The confusion training, validation, and testing matrices.

The step after is the using of classification and regression tree (C&R), that applies to the same mentioned input and target variables. The tree has carried out with highest depth 5, maximum surrogates 5, and using Gini to measure the impurity. As in the previous, the tree has trained 128 as an 80% of all samples, and 16 samples in validation and testing (i.e. 10% for each one). Actually, the tree has achieved 95.83% in 115 of right predictions (agreed with the target) in the training samples, whereas in the testing samples (cells) achieved 92.86% in 13 of correct agreed with the proposed algorithm target by about 100% in validation. Moreover, the most important and effective determined input variable is the Eccentricity by about 0.6786; the Area variable is coming after by about 0.2576, afterwards Perimeter variable by about 0.0358, and the Convex Area variable at the end by about 0.0279. Therefore, by C&R tree, the diagnosis may depend only on the Eccentricity and then Area variables to distinguish the benign cells from distorted ones.

Furthermore, the accuracy, sensitivity, and specificity in training samples have achieved 89.8%, 100%, and 82.1%, respectively. On the other hand, in the test samples, the accuracy achieved 81.25%, sensitivity has 100%, and the specificity got about 66.7%. Although the C&R tree is easier than NN by applying on those exported data, it has achieved only about 92.9%, while NN has achieved 96.9%. Clearly, the previous applications of NN and C&R tree on the exported data is tending to the decision that NN is preferred and more effective than an C&R tree in prediction on these data. In other words, NN is helped to test and check the efficiency of the data resulted from the proposed algorithm in diagnosing and detection the Sickle and all distorted cells. Finally, Matlab 2013a has been used to build the algorithm on Windows 7 with processor Intel ® Core™2Duo CPU T5550@ 1.83GHz and 2.50 GB RAM with 32-bit Operating system. All the images of blood cell were digitized by the optical Nikon microscope.

8 Conclusions

Microscopic image analysis of human blood cells helps as a diagnostic tool for the infected blood cells' detection.

Sickle-cell anemia is one of the most important common types of anemia disease. This paper has been presented a proposed algorithm that can be detecting and counting the Sickle and all distorted cells in a microscopic colored image; even if they are hidden or overlapped. The algorithm has been used circular Hough transform to detect the benign and distorted blood cells. On the other hand, the exported variables data (Areas, Convex Areas, Eccentricity, and Perimeter) for all detected cells (benign, and distorted) have been classified as input variables, whereas a solidity measure variable for all of cells has been constructed as a target variable. In the next step, the neural network and regression tree have been applied to get the right decision for diagnoses and check the effectiveness of the proposed algorithm in detection. The experimental results have been demonstrated high accuracies and success of these models in predicting the infected cells that are contained Sickle Anemia or distorted cells. The performances have been calculated by three statistical measures; classification accuracy, sensitivity, and specificity. This algorithm has been achieved in segmentation and classification processes by about 99.98% out of all input cells are detected, which may have contributed to improve diagnosis of Sickle Anemia diseases. The experimental results have been shown that the effectiveness reaches to 96.9% in the case of applying Neural network and 92.9% when using C&R tree. Therefore, the proposed algorithm is very effective in the detection of benign and distorted red blood cells, additionally that the neural network is more efficient than a C&R tree in the case of testing the quality of the algorithm in detection.

9 References

[1]. B. Chaudhuri and U. Bhattacharya." Efficient training and improved performance of multilayer perceptron in pattern classification." Neuro computing, 34, pp11–27, September 2000.

[2]. B. R. Devi, K. N. Rao, S. P. Setty and M. N. Rao. Disaster Prediction System Using IBM SPSS Data Mining Tool. International Journal of Engineering Trends and Technology (IJETT)-Volume4 Issue8-August 2013 ISSN Volume: 2231.

[3]. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. Classification and Regression Trees, Belmont, California: Wadsworth, Inc, 1998.

[4]. Freeman WJ. "The physiology of perception." University of California, Berkeley. SCI Am. 1991 Feb;264(2):78-85.

[5]. H. A. Khan and G. M. Maruf. Counting clustered cells using distance mapping. International Conference on Informatics, Electronics & Vision (ICIEV), P:1-6, 17-18 May 2013.

[6]. H. Fleyeh, R. Biswas and E. Davami. Traffic sign detection based on AdaBoost color segmentation and SVM classification EUROCON conference, IEEE. 2005-2010. 1-4 July 2013.

[7]. Hany. A. Elsalamony, Alaa. M. Elsayad. Bank Direct Marketing Based on Neural Network. International Journal of Engineering and Advanced Technology IJEAT, ISSN: 2249–8958, Vol-2, Issue-6, Aug 2013.

[8]. Image Processing Toolbox, is available through MATLAB's help menu, or online at: http://www.mathworks.com/help/images/index.html

[9]. K. Thirusittampalam, M. J. Hossain, O. Ghita, and P. F. Whelan. A Novel Framework for Cellular Tracking and Mitosis Detection in Dense Phase Contrast Microscopy Images. IEEE Journal of Biomedical and Health Informatics, Vol. 17, No. 3, MAY 2013.

[10]. M. C. Mushabe, R. Dendere and T. S. Douglas. Automated detection of malaria in Giemsa-stained thin blood smears. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC, P: 3698-3701, 3-7 July 2013.

[11]. S. D. Brown and A. J. Myles. Decision Tree Modeling in Classification. Book chapter Editor: S. D. Brown, R. Tauler and B. Walczak, Book Title: Comprehensive Chemometrics. Elsevier. Pages: 541-569. 2009.

[12]. Stockman GC, Agrawala AK. Equivalence of Hough curve detection to template matching. Communications of the ACM. 1977;20:820-2. 20:820-2.

[13]. T. J. Atherton and D. J. Kerbyson. Size invariant circle detection. Image and Vision computing. Volume: 17. Issue: 11 1999. Pages: 795-803.

[14]. TIAN YuBo, ZHANG XiaoQiu, and ZHU RenJie. "Design of Waveguide Matched Load Based on Multilayer Perceptron Neural Network." Proceedings of ISAP, Niigata, Japan 2007.

[15]. Tom M. Mitchell. "Machine Learning." Learning." Copyright c 2005-2010. Second edition of the textbook. Ch 1. January 19, 2010 McGraw Hill. 119 ACM 2005, ISBN 1-59593-180-5.

[16]. U. S. Department of Health & Human Services, National Institute of Health. http://www.nhlbi.nih.gov/health/health-pics/topics/sca/

[17]. Wikipedia, the free encyclopedia, (Redirected from Neural network). http://en.wikipedia.org/wiki/Neural_network#History of the neural network analogy