The Discrimination of Red Blood Cells Infected by Hereditary Hemolytic Anemia

Ghayda A.A. Al-Talib   Hesham H. Mohameed

ghaydabdulaziz@uomosul.edu.iq   *hesham@ntu.edu.iq*
College of Computer Sciences and Mathematics
University of Mosul, Mosul, Iraq

ABSTRACT

This paper presents a medical application based on digital image processing and Artificial Neural Network (ANN), which can recognize three types of Hereditary Hemolytic Anemia (HHA) that affect the Red Blood Cells (RBCs) and change their shape. Three Feed Forward Back Propagation Learning (FFBBL) Neural Networks are used in hierarchical approach to achieve this goal. The essence of this research is to segment each Red Blood Cell in a separate image and then extract some interesting features from each image in order to present them to the neural networks. The latter will, in turn, take the decision whether the RBC is infected or not. The results showed a recognition rate 92.38 %.

Keywords: Hereditary Hemolytic Anemia, Red Blood Cells, NN

تمييز الكريات الدم الحمراء المتأثرة بفقر الدم الانحلالي الوراثي

غيداء عبد العزيز مجيد الطالب   **هشام هاشم محمد**

كلية علوم الحاسوب والرياضيات، جامعة الموصل

الملخص

يقدم هذا البحث تطبيقاً طبياً يستند إلى معالجة الصور الرقمية باستخدام الشبكات العصبية الاصطناعية، يستطيع التطبيق تمييز ثلاثة أنواع من فقر الدم الانحلالي الوراثي الذي يؤثر على كريات الدم الحمر و يغير شكلها الخارجي. ثلاث شبكات عصبية اصطناعية مرتبة بشكل هرمي استخدمت لتحقيق ذلك الهدف. جوهر العمل هو قطع كل كرية دم حمراء بصورة منفصلة من الصورة الأصلية ثم استخلاص بعض الصفات المفيدة من كل صورة مقتطعة لغرض تقديمهم للشبكات العصبية. و التي بدورها، تتخذ القرار فيما إذا كانت كرية الدم الحمراء مصابة أم لا. أظهرت النتائج دقة في التمييز بلغت 92.38%.

الكلمات المفتاحية: فقر الدم الانحلالي الوراثي، كريات الدم الحمراء، الشبكات العصبية

1. Introduction

Many real world complex problems can be solved through Artificial Intelligence (AI) such as medical, commercial, industrial, and agricultural problems. The most immense area in the AI is the ANN which is widely used to solve complex problems in the pattern recognition, data mining, data security, time series prediction and many problems in various fields. This paper proposes a preprocessing that segment each RBC in separated image, then present each segmented RBC to three ANN arranged in hierarchical way. The ANNs determined whether the RBC infected or not and the type of infection.

The ANN is a computational simulation of biological neural network. ANN successfully simulate the functions of human brain in the term of learning, classification, optimization, prediction, clustering , and generalization[1]. ANN consists of multiple, highly connected neurons that are arranged in single or multiple layers,

depending on the type of network These neurons connect with each other via weights; the type of application is the determining element of these connections and their weights[2].

A three feed forward back propagation learning neural networks in hierarchical way is implemented to produce a system which is able to recognize three types of hemolytic anemia that affect RBCs and change their shape.

2. Related Work

In the medical field, there are many related works for using medical image processing to discriminate the hematological diseases.

In 2008 Basim Alhadidi and Hussam Nawwaf Fakhouri proposed an iron deficiency anemia blue and red cells calculating system. They implemented an algorithm that achieves an automated way for the analysis of images taken for intestine villi. This algorithm will count the number blue and the red stained cells blood cells that contain iron in each villi alone. And also calculate the percentage of blue cells and red cells in the image[3].

In 2009 Hirimutugoda and Wijayarathna presented a research about using artificial intelligence for determining hematologic diseases, namely Malaria, thalassemia, and possible other abnormal red cell, they reached to 86.54% successfully recognition[4].

Also, in 2009 Makkapati and Rao presented a scheme based on HSV color space to segment RBCs and Malaria parasites by detecting dominant hue range and calculating optimal saturation thresholds, they reached to 83.54% sensitivity and 98% specificity [5].

Furthermore, Kondo and Ueno presented in 2011 a medical image diagnosis system for lung cancer detection by Revised GMDH-type neural network that uses heuristic self-organization network[6].

3. Background

Anemia can be defined as low hemoglobin concentration and is appeared in several types. One of these types is Hereditary Hemolytic Anemia which occurs when the body destroys the abnormal RBC's faster than the bone marrow can create new normal RBC's[7]. Hereditary Hemolytic Anemia occurs as a result of one of three causes[8]:

- Enzymatic Defect: It occurs when one of the Enzymes inside the RBC is defected. It results into two hemolytic anemia, (Pyruvate Kinase deficiency anemia) and (Glucose -6- Phosphate Dehydrogenase deficiency anemia). The former doesn't change the outer shape of the RBC. It needs some biological tests and the infection does not appear in the image of the blood smears, so it's out of the scope of this research. On the other hand, the latter is an anemia which is resulted from the deficiency of cellular enzyme called G6PD. In this case, when some free radicals and oxygen enter the RBC, they oxidase the DNA of the RBC and result in formation of bodies that stick on the wall of the RBC and then causes the membrane of RBC to be broken at that side[9] see (Fig.1a) and (Fig.1b).
- Membrane Defect: It occurs when one of the membrane proteins is defected. It results into two hemolytic anemia, (Spherocytosis anemia) and (Elliptocytosis anemia). The former changes the color of the RBC, not the shape, so it's out of the scope of this research, while the latter, which is usually called Elliptocytosis or

oval, is an anemia which changes the RBC's shape to Ellipse[9] (see (Fig.1c) and (Fig.1d)).

- Hemoglobin Defect: It occurs when the hemoglobin is defected. It results into two hemolytic anemia (Thalassemia) and (Sickle Cell Anemia). The former is divided into four types; all of them do not change the shape of the RBC, while the latter results from substituting a single amino acid passed in the globin chain of the RBC's hemoglobin by another one which is regarded as abnormal. When this type of RBCs is put in a situation in which there is a decrease in the oxygen, the RBC's will consequently undergo sickling. This shape makes the RBC fragile and sticky; which leads to decomposition of the cell[9] ((see Fig. 1e)).

This paper proposes discriminating three types of hereditary hemolytic anemia, these three types are: G6PD deficiency anemia, hereditary elliptocytosis anemia, and Sickle cell anemia.
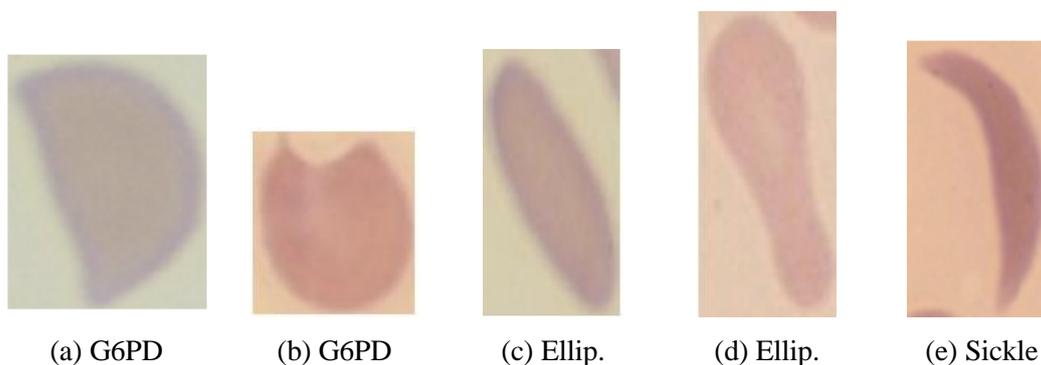

| (a) G6PD | (b) G6PD | (c) Ellip. | (d) Ellip. | (e) Sickle |

Figure (1). Some of Hereditary Hemolytic Anemia Types

4. Proposed Scheme

Each RBC has been segmented in separated images, then some interesting features were taken from each image and presented to three Feed Forward Back Propagation Learning Neural Networks arranged in hierarchical approach. So the pre-processing for segmenting each RBC will be presented first, and then the proposed neural networks are presented.

4.1 Pre-Processing:

Though, it differs from one research to another, but all the pattern recognition researches must contain this step. The pre-processing in this research consists of six steps where the output from each step feeds as input to the next step. These steps are:

1. Converting the input image from Red Green Blue RGB color space to YCbCr color space. The new color space with three components Y, Cb, Cr which are given as follows[10]

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.996 \\ -37.797 & -74.203 & 112.00 \\ 112.00 & -93.786 & -18.214 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The image is converted to this color space to take advantage of using the luminous component Y, which feeds as input to the second step.

2. Applying the k-means clustering algorithm to the input image (i.e. Y component), taking into consideration that there are two centers in the processed image. This makes sense because considering two centers will make the RBCs, White Blood

Cells WBC's, platelet, and artifacts as a class and the background as a second class, ((see Fig. 2)). The output from this step (i.e. k-means result) feeds to the third step[7].
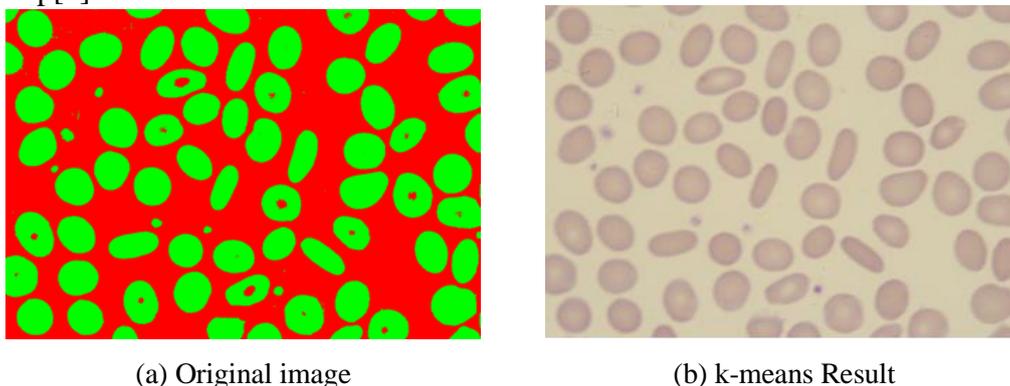

| (a) Original image | (b) k-means Result |

Figure (2). Clustering Result

3. Applying Canny edge detector[11] to the gray image resulted from converting the colored k-means result (Fig.2 b) to gray. Applying the edge detector will obtain only the edges of the objects inside the image. The result is a black and white image (BW image) that contains RBC's edges, WBC's edges, platelet edges and artifacts edges.

4. This step performs two operations: the first one is eliminating the clipped cells that are located on the perimeter of the image, since these cells are sometimes hard to analyze even with the pathological analyzer. This is done by eliminating each open object. The second operation is filling all the enclosed objects inside the image[12], that is the RBC's, WBC's, Platelets, and Artifacts (see Fig. 3a). Then, applying the canny edge detector to the filled image. This step is necessary because the normal RBC may contain hole inside it in the clustered image (see Fig.2 b). This hole will be eliminated in the filled image (see Fig.3 a). It is worth noting that only RBC's objects are needed, so the other elements will be eliminated in the next step.
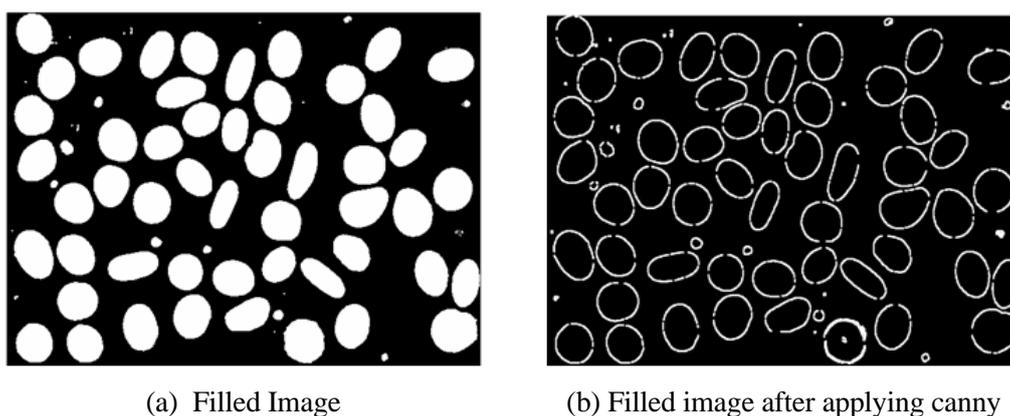

| (a) Filled Image | (b) Filled image after applying canny |

Figure (3). Producing Enclosed Edges Only

5. This step will eliminate the WBC's, platelets and the artifacts as they are out of the scope of this work. This is done through drawing the boundary pixels of each enclosed object in a separate image if the pixels of the boundary are in between 150 and 1000 pixels. Otherwise, ignoring the values out of this range.

Ignoring every enclosed object that either has less than or equals to 150 boundary pixels will eliminate the platelet, or eliminate the artifacts. While, ignoring every enclosed object that has greater than or equals to 1000 boundary pixels will eliminate the WBC, since these objects are enclosed, but they are not RBCs. The result of this step is the segmented black and white RBC's which will be presented to the last step of pre-processing (see Fig.4).
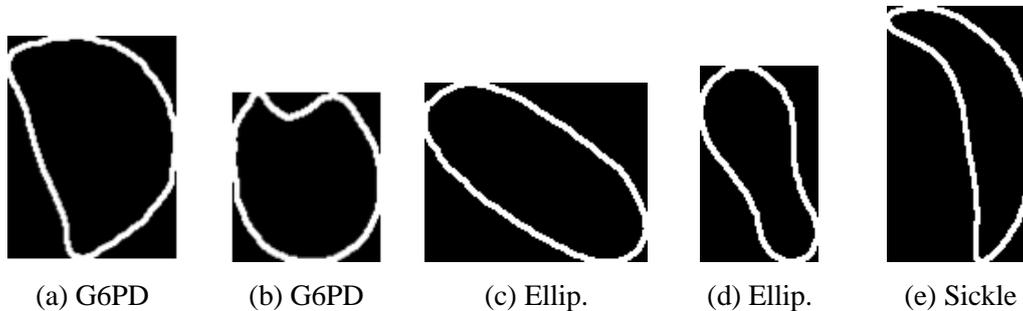

| (a) G6PD | (b) G6PD | (c) Ellip. | (d) Ellip. | (e) Sickle |

Figure (4). Segmented black and white RBC's

6. Though most of the segmented RBCs are now ready to be presented to the neural network, but this step is very important because there are occasions on which the segmented rectangle of the RBC has some noise from the neighboring cell(s), The noise is surrounded by red circle in (Fig. 5a) and (Fig.5b), when some parts of that cell fall inside the segmented RBC rectangle. This step eliminates that noise by deleting every object having less than 100 pixels inside the segmented RBC image, (see Fig5.c). After deleting the noise from all the segmented images that is, segmented RBC's, they are ready to be presented to the ANN.
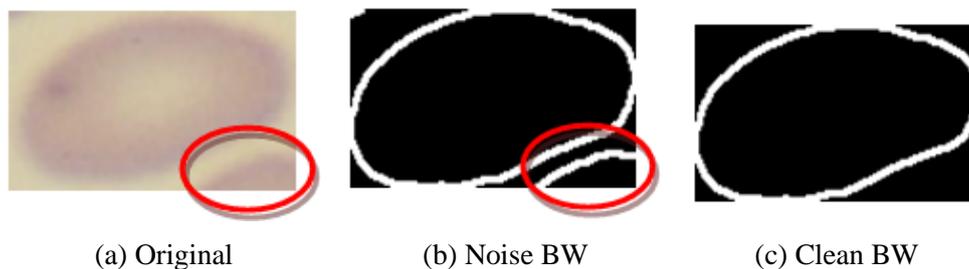

| (a) Original | (b) Noise BW | (c) Clean BW |

Figure (5). Deleting Noise

4.2 Artificial Neural Networks:

ANN is a set of simple units called Processing Elements (PEs). These processing elements are arranged in layers and connected to other processing elements through weights in order to form the neural network. The processing element itself performs some simple computation such as computing the weighted sum of its inputs and then testing some activation function to produce the output which, in turn, passed to the next layer[13].

There are two types of ANN supervised and unsupervised. This paper proposes using three supervised ANN. In supervised ANN, the input presented to the input layer, and the weights are adjusted depending on a comparison of the network output and the target (see Fig.6), until network output and target are matched[14]. The essence of this work is to extract some descriptors from each cell and present these descriptors to the ANN.
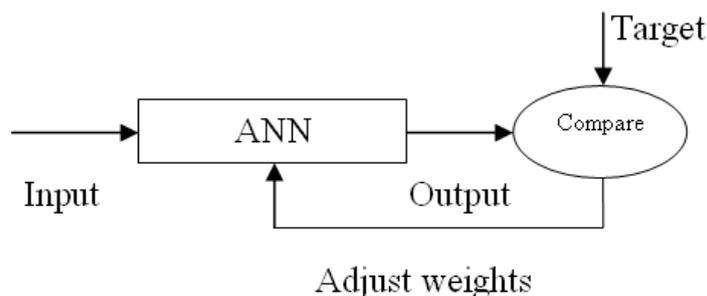
Figure (6). Supervised learning

As it is previously mentioned, three types of hereditary hemolytic anemia were discriminated by hierarchical neural networks. It might be thought that the network must be trained on four types only, which are G6PD deficiency anemia, Elliptocytosis anemia, Sickle cell anemia, and the normal, or not infected, RBC. Rather, the Fieldwork has some complication. For example, the cell infected with G6PD deficiency may appear in two shapes: the first one cuts off a large part (perhaps half of the RBC), (see Fig.1a or Fig.2a), while the second cuts off a small part of the RBC, (see Fig.1b or Fig.2b).

Likewise, the RBC infected by Elliptocytosis Anemia may appear in two shapes: the first one is fully ellipse, (see Fig.1c or Fig.2c), while the second has an ellipse shape, but thinner from one of its ends, (see Fig.1d or Fig.2d).

Each one of the two types of G6PD deficiency will be handled by separated neural network, while the two forms of elliptocytosis anemia used in the training treated as separated type, i.e. they were given different output codes during the training of the network. Then, the codes are gathered when given the result of infection.

One more complication occurs when two cells appear attached in the image (see Fig.7). This kind of cells is regarded as normal. Though, there is a possibility that one of these cells is infected, but this possibility is low.

Figure (7). Connected cells appear in the samples

As mentioned previously, three feed forward back propagation learning neural networks are used in hierarchical way, (see Fig.8), to achieve the recognition. Each segmented RBC is presented to the first neural network (NET 1) to check whether the RBC has circular shape or not. Depending on the result of (NET 1), the segmented RBC is sent either to the second neural network (NET 2), circular, or to the third neural network (NET 3), not circular.

For better understanding of the networks and their functions each network will be presented separately:
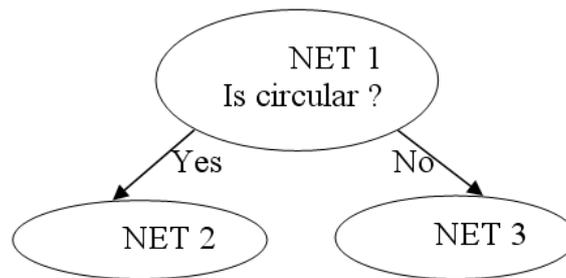
Figure (8). Proposed approach

1. NET 1: This network trained on the following three features:
*   Compactness: The compactness for a 2D shape is defined as the ratio between its perimeter and area. The compactness is equal (perimeter2)/area[15].
*   Eccentricity: To find the eccentricity of an object, two lines must be drawn: the first one connects the farthest points on the border of the object, and the second one should pass vertically the first one and also connect the farthest points on the border of the object[16]. These lines can be used as descriptors. The eccentricity can be defined as the ratio between the longest line to the shortest one. In the next figure (Fig.9) the eccentricity is the ratio between the lines A and B i.e. Eccentricity = A/B.
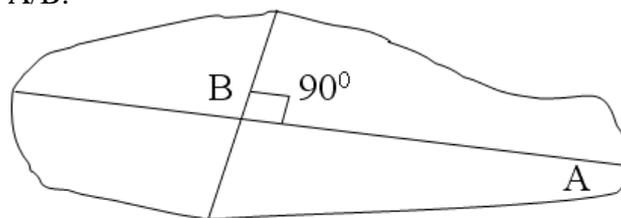

Figure (9). Eccentricity

*   Elongatedness: For each object a bounding rectangle can be drawn. The minimum bounding rectangle, dotted rectangle in (Fig. 10), can be used as a descriptor. The elongatedness is the ratio between the longest side of the minimum bounding rectangle to the shortest one[13]. In the (Fig. 10) the elongatedness is the ratio between A and B i.e. Eccentricity = A/B.
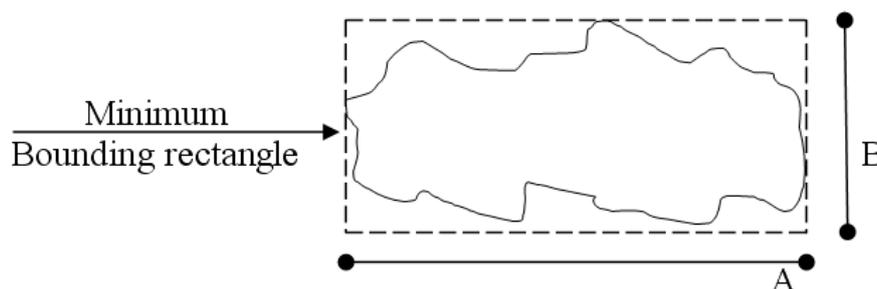

Figure (10). Elongatedness

The network NET 1 has three layers; its topology is illustrated in (Fig. 11). The input layer has three neurons, the hidden layer has five neurons, while the output layer has single neuron. It takes 1 second to reach stability.
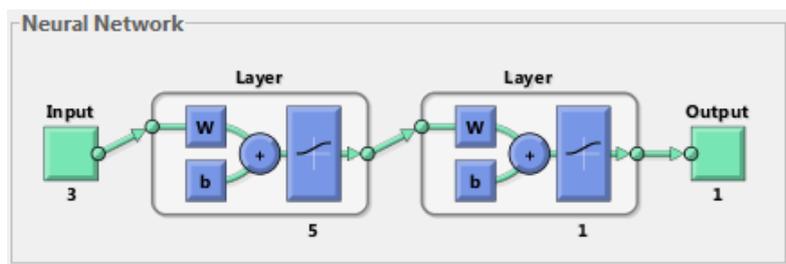
71

Figure (11). NET 1 topology

This network reaches stability after 112 epochs when the mean squared error was 1*10E-6 (see Fig. 12).

Each segmented RBC is presented to NET 1, the three features are calculated to the RBC, and then it is presented to the network to determine whether the RBC have circular shape. If NET 1 responds by output 1, i.e. NET 1 decides that the RBC has circular shape, then the RBC is either infected by G6PD deficiency anemia that cuts off small part of the RBC (Fig.1 b), or it is normal RBC. It is note-worthy that in both cases, the RBC has circular shape. The RBC, then is sent to NET 2 to determine the correct type.
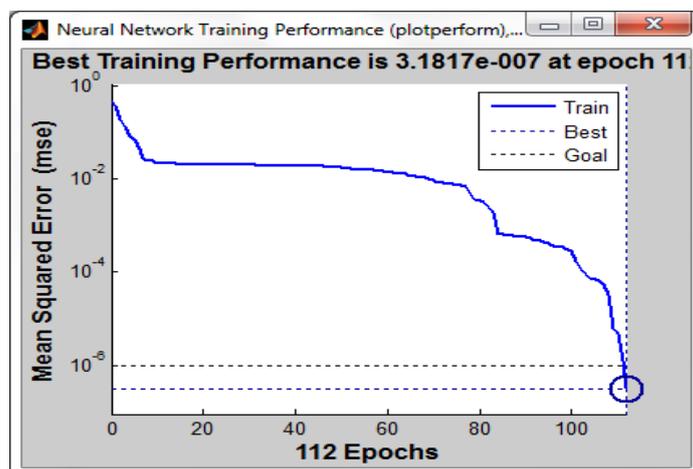

Figure (12). NET 1 performance

2. NET 2: This network trained on one feature, that is the signature.
- Signature: It is a simple descriptor that describes the boundary of the shape. The signature can be calculated through creating a vector of 360 points; each point represents the distance from the center of the object to the boundary of it at that angle. The signature can be expressed as follows:

$$\text{Signature } (\Theta) = \sqrt{(X_c - X_\Theta)^2 + (Y_c - Y_\Theta)^2}$$

Where X_c= X coordinates at the center, Y_c=Y coordinates at the center, X_Θ = X coordinates of the boundary point at angle Θ, Y_Θ = Y coordinates of the boundary point at angle Θ, Signature (Θ) = signature value, distance, at angle (Θ). (see Fig. 13)
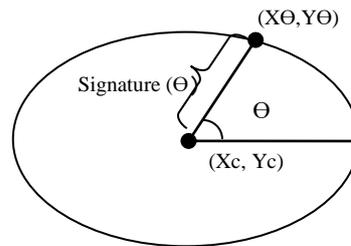
Figure (13). Signature

Before using the signature, it is made as a scale invariant through dividing all the values of the signature by the maximum value of it. This will make all the values of the signature between the interval (0,1] and [13]. Finally, to reduce the dimensionality of the scale invariant signature, only 57 points of it have been taken as a descriptor, these 57 points starting at theta=1,7,13 ……343.

The network NET 2 has four layers and its topology is illustrated in (Fig. 14). The input layer has 57 neurons, the first hidden layer has 57 neurons, the second hidden layer has 30 neurons, the output layer has single neuron. It takes 96 epochs to reach stability after training for 16 minutes and 57 seconds when the mean squared error was $1*10E-6$ (see Fig. 15).
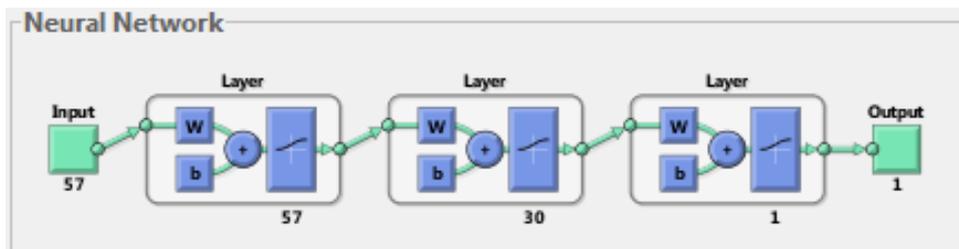

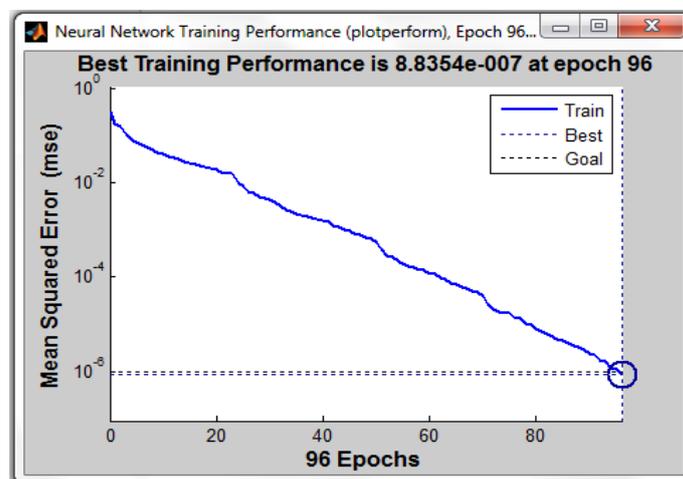**Figure (14) .** NET 2 Topology

Figure (15). NET 2 performance

If the RBC is presented to this network and the network responds by 1 at the output layer, then the presented RBC is infected by G6PD deficiency that cuts off small part of the RBC, else the RBC is normal.

3. NET 3: This network is trained on one feature, that is the signature. The network NET 3 has four layers and its topology is illustrated in (Fig. 16). The input layer has 57 neurons, the First hidden layers have 57 neurons, the second hidden layer has 30 neurons, and the output layer has 6 neurons. It takes 120 epochs to reach stability

after training for 24 minutes, when the means squared error was 1*10E-5, (see Fig. 17).
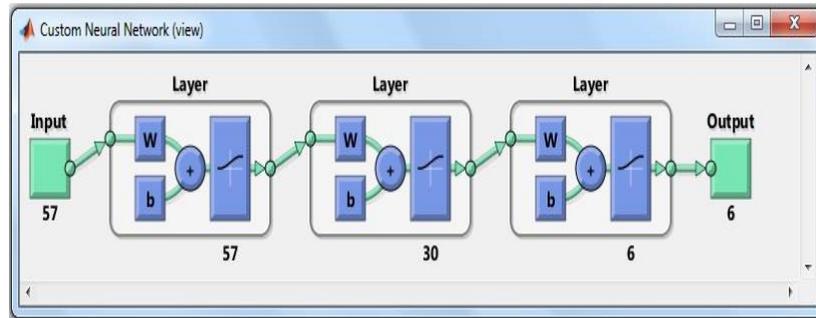

Figure (16). NET 3 Topology

When the RBC is presented to NET 1 and NET 1 responds by 0 at the output layer, then it means that the RBC does not have circular shape. In this case, the RBC is sent to network NET 3 which, in turn, checks the type of the RBC, that can be either Sickle cell anemia, or G6PD anemia which cuts off large part of the RBC, or one of the two forms of Elliptocytosis anemia (Fig. 1c) and (Fig. 1d), or two cells connecting together, (see Fig. 7). Again the latter type is regarded as normal; all the other four types are infections.
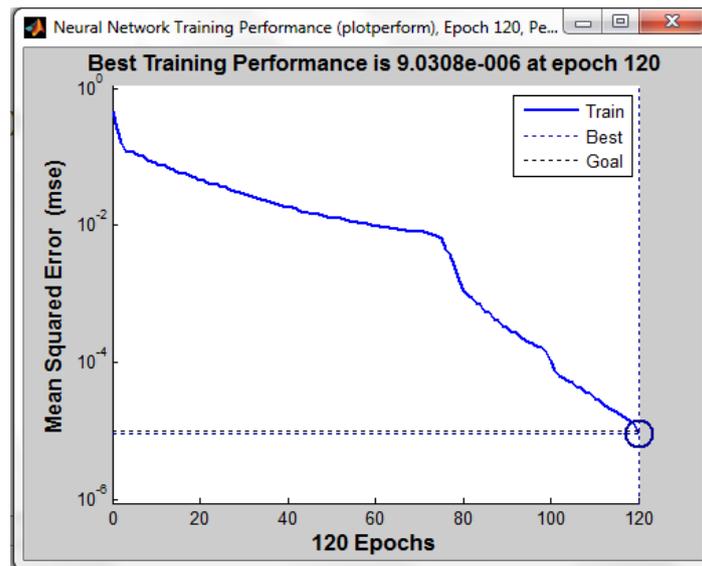

Figure (17). NET 2 performance

5. Results

The data set used in this paper is 407 cells. The cells are taken from ten images and the data set is presented as shown in the following table:

Table 1. The used data set

Image No.	No. of Processed Cells	No. of Infected Cells	Type of Infection
1	**26**	**12**	**Elliptocytosis Anemia**
2	**52**	**16**	**Elliptocytosis Anemia**
3	**72**	**44**	**Elliptocytosis Anemia**
4	**39**	**2**	**G6PD Anemia**
5	**54**	**6**	**G6PD Anemia**
6	**27**	**4**	**G6PD Anemia**
7	**45**	**6**	**G6PD Anemia**
8	**23**	**6**	**Sickle Cell Anemia**
9	**22**	**7**	**Sickle Cell Anemia**
10	**47**	**0**	**Normal**
∑	**407**	**103**	

For better understanding of networks results, the results of each network will be discussed separately and a presentation of the results of the all structure will be done:

1. NET 1: The Network NET1 has been trained on 90 cells, i.e. (22.11 %) of the data and is tested on the all 407 cells. The training database contains 40 non-circular cells and 50 circular ones. The next three figures, particularly (Fig.18), (Fig.19) and (Fig.20) show the compactness, Eccentricity and Elongatedness training data respectively.
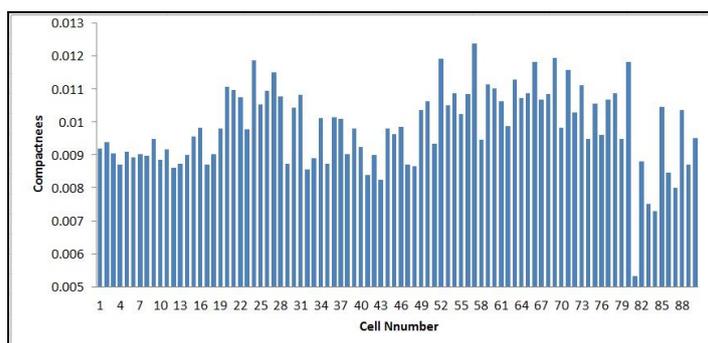

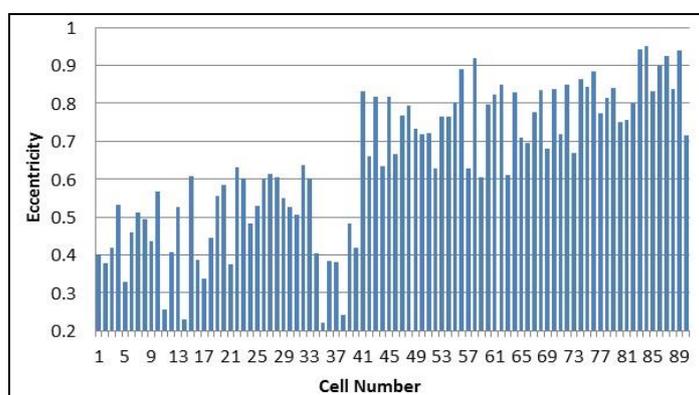**Figure (18).** NET 1 compactness training data

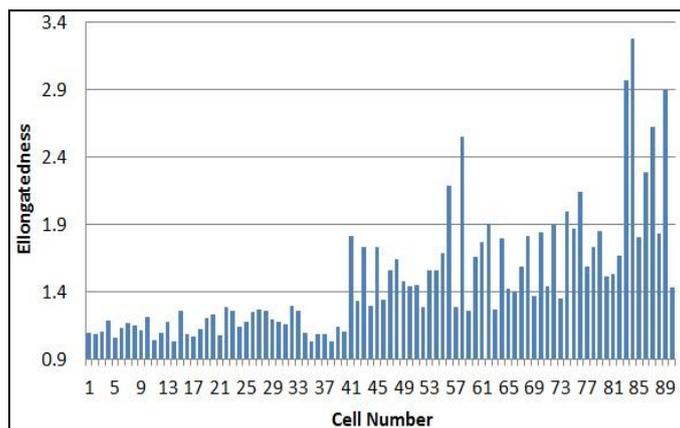**Figure (19).** NET 1 Eccentricity training data

Figure (20). NET 1 Elongatedness training data

NET 1 shows that (98.53 %) of the cells had been recognized correctly. The following table shows NET 1 Results.

Table 2. NET 1 Result

Image No.	No. of Circular Cells	No. of Not Circular Cells	No. of Cells Not Recognized Corectly By NET 1	Percentage %
1	12	14	0	0
2	37	15	0	0
3	21	51	1	1.38
4	36	3	0	0
5	46	8	1	1.58
6	26	1	2	7.4
7	43	2	0	0
8	15	8	2	8.69
9	13	9	0	0
10	46	1	0	0
∑	295	112	6	1.47

2. NET 2: The Network NET2 has been trained on 22 cells, i.e. (7.45 %) of the 295 circular cells and is tested on all of them. The training database contains 11 cells infected by G6PD and 11 normal cells. The network NET 2 shows that (94.58 %) of the cells has been recognized correctly. Table 3 shows NET 2 Results.

Table 3. NET 2 Result

Image No.	No. of Circular Cells	No. of Cells Not Recognized Corectly By NET 2	Percentage %
1	12	2	16.6
2	37	2	5.4
3	21	2	11.11
4	36	2	2.77
5	46	2	8.69
6	26	1	3.84
7	43	2	4.65
8	15	3	20
9	13	0	0
10	46	0	0%
∑	295	16	5.42

3. NET 3: The Network NET3 has been trained on 33 cells, i.e. (29.46 %) of the 112 non-circular cells and is tested on all of them. The training database contains 8 cells which were originally 16, but are connected together to be 8 (Fig. 7), and 6 cells of full ellipse shape elliptocytosis anemia 15 cells have elliptocytosis anemia that is thin from one of its parts, 3 have sickle cell anemia, and 1 is infected with G6PD deficiency anemia. The network NET 3 shows that (90.18 %) of the cells have been recognized correctly. Table 4 mhows NET 3 Results.

The final results show that 92.38 % of the cells recognized correctly.

Table 4. NET 3 Result

Image No.	No. of Not Circular Cells	No. of Cells Not Recognized Corectly By NET3	Percentage %
1	14	0	0
2	15	1	6.66
3	51	0	0
4	3	1	33.33
5	8	1	12.5
6	1	0	0
7	2	1	50
8	8	0	0
9	9	7	77
10	1	0	0
∑	112	11	9.82

6. Conclusion

It can be noticed that network (NET 1) has the largest ratio of succession (98.54 %). This is because the training data of NET 1 are very different from each other in the properties. It is worth-mentioning that the first 40 cells have small eccentricity and small elongatedness compared with the remaining 50 cells; (see Fig.19 and Fig.20).

The compactness does not have difference for all the 90 cells, but it is important to be added to the feature, so that the two forms of the G6PD deficiency can be separated to the two different networks, NET 2 and NET 3.

The final result shows that the resulted accuracy of overall approach is (92.38%), which is the mean of NET 2 and NET 3 results. NET 1 error doesn't directly affect the recognition because NET 1 job is to distinguish the circular and not circular RBC's. Though NET 1 has 1.47% error, but this error affects the recognition in accumulative manner.

7. Future Work

This work can be extended to recognize all the Hereditary Hemolytic Anemia types that do not need biologist analysis. This can be done by analyzing RBC texture and discovering, through color changing, the type of infection.

8. Acknowledgements

REFERENCES

[1] H. Uğuz, 2010, "A Biomedical System Based on Artificial Neural Network and Principal Component Analysis for Diagnosis of the Heart Valve Diseases," *Springer,* pp. 62-72.

[2] Q.K. Al-Shayea, 2011, "Artificial Neural Networks in Medical Diagnosis," *IJCSI,* vol. 8, pp. 150-154.

[3] B. Alhadidi and H. N. Fakhouri, 2008, "Automation of Iron Difficiency Anemia Blue and Red Cell Number Calculating by Intictinal villi Tissue Slide Images Enhancing and Processing," *2008 International Conference on Computer Science and Information Technology,* pp. 407-410.

[4] Y.M. Hirimutugoda and G. Wijayarathna*,* 2009, "Artificial Intelligence-Based Approach for Determination of Haematalogic Diseases," *IEEE*.

[5] V.V. Makkapati and R. M. Rao, 2009, "Segmentation Of Malaria Parasites In Peripheral Blood Smear Images" *IEEE,* pp. 1361- 1364.

[6] T. Kondo and J. Ueno, 2011, "Medical Image Diagnosis of Lung Cancer by Revised GMDH-type Neural Network Using Heuristic Self-Organization," *SICE Annual Conference,* pp. 1254- 1259.

[7] Z.S. Zhang, X.-M. Wang, Z.-P. Han, L. Yin, M. X. Zhao, and S.-C. Yu, 2011, "Physicochemical properties and inhibition effect on iron deficiency anemia of a novel polysaccharide iron complex," *Elsevier,* p. 489 492.

[8] J.O. Armitage*,* 2008, "Atlas of Clinical Hematology," *Springer*.

[9] M. Longmore, L. B. Wilkinson, E. H. Davidson, A. Foulkes, and A. R. Mafi, 2010, "Oxford Handbook of Clinical Medicine".

[10] N. K. Patil, R. M. Yadahalli, and J. Pujari, 2011, "Comparison between HSV and YCbCr Color Model Color-Texture based Classification of the Food Grains," *International Journal of Computer Applications,* vol. 34, pp. 51-57.

[11] G.H. Kumar and T. Jipeng, 2012, "Different Edge Detection Algorithms Comparison and Analysis on Handwritten Chinese Character Recognition," *International Journal of Computer Applications,* vol. 47, pp. 20-23.

[12] S. Sharma*,* 2010, "Digital Image Processing," *S.K. kataria & Sons*.

[13] K.U. Rani, 2011, "Analysis of Heart Diseases Dataset using Neural Network Approach".

[14] S. Chidrawar, S. Bhaskarwar, and S. Chidrawar, 2012, "Detection of Brain Tumor in Radiographic Images using Neural Network," *IJCA,* pp. 4-5.

[15] E. Bribiesca, 2007, "An easy measure of compactness for 2D and 3D shapes," *Elsevier,* pp. 543-554.

[16] R.C. Gonzalez and R. E. Woods, 2008, "Digital Image Procesing".