

Detection of anaemia disease in human red blood cells using cell signature, neural networks and SVM

Hany A. Elsalamony^{1,2}

Received: 12 October 2016 / Revised: 1 July 2017 / Accepted: 8 August 2017 /

Published online: 19 August 2017

© Springer Science+Business Media, LLC 2017

Abstract Anaemia disease attacks and deforms the circular red blood cells. Latterly, it has classified as a very dangerous disease. Many papers have been presented approaches for tracking and detection of anaemia cells before; but this time, elliptocytosis, sickle, and burr cells have detected based on their shape signatures. In tested images, some of the cells have been formed unknown shapes resulted from stuck operation in the instant image capture. This shape have not been belonged to any of anaemia kinds, then they have considered as cells with unknown shape. Here, the using of Circular Hough transforms, watershed segmentation and some of the morphological methods has been urgent manner to enhancing and preparing tested images. The performance of proposed algorithm have been achieved highly accuracy by testing 45 colourful microscopic images in 15 samples from patients already have anaemia disease. The Support Vector Machine (SVM), back propagation (BP) and self-organising map (SOM) neural networks have been applied on all information data of mentioned kinds of anaemia.

Keywords Elliptocytosis · Sickle cells · Burr cells · Cell signature · Circular Hough transforms · Watershed segmentation · SOM neural network · BP neural network · SVM

1 Introduction

Red blood cells are being inside the huge bones of the body in the spongy marrow. Along the time, the main action of marrow is the renewing and replacing old red blood cells. Approximately, 120 days is the normal life for RBCs in the stream of human blood. Their main role in this short life is to carry oxygen to all body and remove carbon dioxide (a waste product) from

✉ Hany A. Elsalamony
h_salamony@yahoo.com

¹ Mathematics Department, Faculty of Science, Helwan University, Cairo, Egypt

² Computer Science Department, Arts & Science College, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

it. The normal shape of RBCs is like a disc that gives a chance of moving easily through the blood vessels [11, 22]. Anaemia disease is working on deformation of RBCs into various other shapes that causes non-attendance of iron in the blood. Iron is essential to the human body for the creation an iron-rich protein called haemoglobin, which assists red blood cells with carrying oxygen from lungs to whatever is left of the body. When the normal number of red cells is lower in the blood or deficiency of haemoglobin, this ailment happens [10, 12, 14, 15, 32].

Echinocyte (Burr) comes from the Greek word meaning “sea urchin,” which relates to its shell-like appearance. The shape of Burr cells have serrated edges over the entire surface of the cell. Burr cells, are reversible, meaning that this change can be the result of the cell’s environment, pH of the medium (including the glass slides on which blood smears are made), the metabolic state of the cell, and the use of some chemical substances. Actually, Burr cells appear due to uremia caused by mild hemolysis of red blood cells as seen in hemolytic anaemia, hypomagnesemia, hypophosphatemia, and pyruvate kinase deficiency. The projections of the cell membrane may be sharp or blunt, are usually numerous, and tend to be evenly spaced around the circumference. Spicules are usually of uniform size with short, evenly spaced spicules and preserved central pallor that is usually artifactual (observed in uremia and liver disease). There has been no clear definition of a reference range for burr cells in normal individuals and among patients with various diseases [3, 4]. Echinocytes were first described in 1957 in association with renal disease and uremia and may be exacerbated during hemodialysis [1, 2]. Echinocytes are commonly found in patients with liver failure, hyperbilirubinemia, vitamin E deficiency, pyruvate kinase deficiency and are seen occasionally as artefacts on the peripheral smear.

One of most famous kind of anaemia is named sickle; this name comes when the disc shape of red blood cells deformed into a crescent shape. The substantial danger of sickle cells is that they obstruct the flow of the stream of blood in the veins, limbs, and organs. The resistance in flowing of blood can cause pain; organ hurt and push the probability of illness. Moreover, the sickle cells usually die after approximately ten to twenty days, and the bone marrow cannot renew red blood cells fast enough to replace that dying RBCs [30]. In addition, this kind of anaemia is most famous in people whose families come from Mediterranean nations, Africa, South, or Central America, particularly Panama, Caribbean islands, Saudi Arabia, and India. The patient number of the individuals that are suffering from sickle cell anaemia in the United States ranged between 70,000 to 100,000 chiefly African-Americans. The primary decision of it is dependent on blood test examination that can recognize sickle cells [12, 28].

Another kind of anaemia called elliptocytosis. It described in 1904 and then recognized as a hereditary condition in 1932. Medically, it is so hard to determine the hereditary elliptocytosis. This kind of anaemia is estimated to affect an average of 60–150 cases per 10,000 people in African and Mediterranean countries, whereas to affect 3–5 cases per 10,000 people in the USA, and between 1500 and 2000 in Malayan per 10,000 people [9, 23].

Actually, the danger and threat of anaemia disease have pushed to work in this paper for giving some assistance in resistance and diagnosis. The extraction of information from microscopic images represents a big challenge. The regularization of images, image segmentation followed by feature extraction and its classification have included in this type of research. The proposed algorithm begins by detecting the normal RBCs and abnormal ones especially burr, sickle, elliptocytosis cells based on their shape signatures on blood smears [13]. Circular Hough Transform (CHT), watershed segmentation, and morphological tools have used through detection process to select and marked the normal cells. In this manner, the

deformed cells have selected from the normal ones and based on cell shape signatures, as in section 5. The deformed cells information and their related data have formed as input variables to be suitable for use in SVM and two kinds of neural networks; self-organising map (SOM) and back-propagation. These variables are Areas, Convex Area, Perimeter, Eccentricity, Solidity, Ratio, absolute deviations, denoted by AVD, and variable of absolute subtractions among input and saved signature values (see Eqs. 6, 7, 10, 11). The neural networks and decision tree have applied to get optimum classification method with the proposed algorithm [7, 16]. The differences in the shape of Normal, Burr, Sickle, and Elliptocytosis cells have shown in Fig. 1.

The remainder of the paper is organised as follows; Section 2 focuses on related studies. An overview on Circular Hough Transform is presented in section 3. Section 4 is introduced an overview of SVM and neural networks. The shape signature of cells is presented in Section 5. The proposed algorithm is discussed in section 6, and the experimental results of the proposed algorithm are detailed in section 7. Section 8 concludes the paper.

2 Related studies

Recently, research using image processing has increased especially on the blood cells detection and the determination of illnesses. In April and June 2016, two papers have presented algorithms for counting and detecting healthy and many of the unhealthy human blood cells on a smear based on circular Hough transform. These cells have detected and classified into normal and different kinds of anaemia like a sickle, microcytic...etc. The neural network has applied on their extracted data to evaluate the algorithm [14, 15]. The experimental results have been demonstrated high accuracy, and the proposed algorithm has achieved the highest detection dependent on CHT and morphological methods only [13]. In this paper, the algorithm has been used shape signature method and some modifications for helping in the detection process.

Yi and others in 2015 proposed a three-dimensional order technique for naturally deciding the morphologically typical RBCs in the image of numerous human RBCs that got by off pivot advanced holographic microscopy (DHM). The RBC 3D images recorded in the beginning by DHM and after that computational numerical algorithm recreated the stage images of numerous RBCs. They chipped away at three common RBC shapes, stomatocyte, discocyte, and echinocyte for training and testing. Nonetheless, the irregular RBC shapes characterised as a fourth class. Actually, they are subject to ten components extracted from every RBC after segmenting the reconstructed phase images by using a watershed transform algorithm, which include projected surface area, average phase value, mean corpuscular haemoglobin, perimeter, mean corpuscular haemoglobin surface density, circularity, mean phase of centre part,

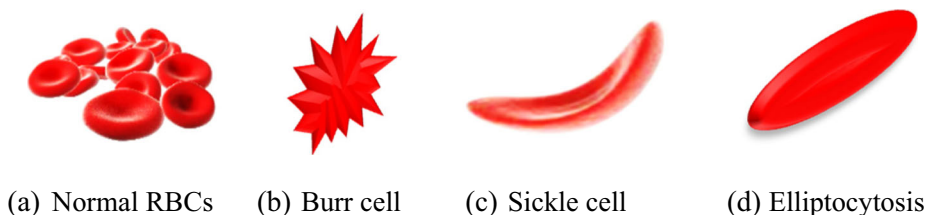


Fig. 1 The differences in shape among normal RBCs

sphericity coefficient, elongation, and pallor. Likewise, they applied the principal component analysis algorithm to diminish the dimension number of variables and built up the Gaussian mixture densities utilising the anticipated information with the initial eight principal components. Thus, the Gaussian mixtures used to plan the discriminant capacities in light of Bayesian decision hypothesis. Their exploratory results demonstrated that the proposed strategy could deliver good results for figuring the rate of each ordinary typical RBC shape in a recreated stage image of various RBCs [33].

In 2014, Lee and Chen introduced a neural network model with a classifier which use the visual data separated from the images of red blood cell to determine if a red blood cell is ordinary or strange. They clustered the visual elements into two primary gatherings, in particular, texture cluster groups and shape cluster groups depending on the element properties. The input feature clusters handled utilising parallel and course construction with multiple input layers. Their trial results indicated a noteworthy change in order precision in the proposed framework when contrasted with the single input layer classifier with late feature selection algorithms [23].

Rashmi Mukherjee in 2014 displayed an assessment of morphometric elements of placental villi and vessels in preeclampsia and typical placentae. The study included light microscopic images of placental tissue areas of 40 preeclampsia and 35 normotensive pregnant women. The villi and vessels depicted from preprocessing and segmentation of these images. She applied principal component analysis (PCA), Fisher's linear discriminant analysis (FLDA), and hierarchical cluster analysis (HCA) to recognize placental (morphometric) highlights, which are basic features in microscopic images. She accomplished five huge morphometric highlights (>90% general discrimination accuracy) recognized by FLDA, and PCA returned three significant principal components cumulatively clarified 98.4% of the total variance [25].

In the same year, Tomari and others showed a proposed PC supported frameworks to mechanize the procedure of discovery and recognizable proof of RBCs from blood smear image. This framework has comprised a blend of three fundamental stages, segmentation and handling, features extraction, and classification. They really were extracted RBCs regions from the background using global threshold method applied on green channel color image. Otsu segmentation method applied on green color channel image with a progression of post handling filter to concentrate strong RBCs shape from the background. In addition, the filter and connected component labeling used to remove noise and holes. According to feature extraction, the fusion of compactness and HU moment invariant were successful to distinguish the normal and abnormal shape with low computational resources and high discriminate power. At last, the ordinary/irregular cells arranged to utilize neural network with four hidden layers. The system tried to classify the RBCs, and the outcomes accomplished 83% accuracy, 82% normal accuracy, and 76% normal review [31].

In July 2013, Mushabe and others presented an algorithm was distinguished and counted red blood cells (RBCs) and parasites to perform a parasitemia estimation. Morphological operations and a histogram-based threshold used to classify red blood cells. They utilized boundary curvature calculations and Delaunay triangulation to the split overlapped red blood cells. A Bayesian classifier with RGB pixel values employed as elements to classify the parasites. The outcomes indicated 98.5% sensitivity and 97.2% specificity for distinguishing tainted red blood cells [26].

In March 2013, Minetti and others have presented an overview of common mistakes using the most popular methodologies in red blood cell research and how to avoid them. In addition, they introduced a number of standards for data comparison between the different labs and

techniques. They were considered flow cytometry, patch-clamp measurements, biochemical analysis, flux measurements and dynamic fluorescence imaging as well as emerging single-cell techniques, such as the use of optical tweezers and atomic force microscopy. They concluded that to clarify the intercellular variability of responses, measurements in cell suspensions should be combined with single-cell techniques such as fluorescent live cell imaging, FCM and/or patch-clamp approaches. Though, even between single-cell techniques, there are regularly differences and confusing interpretations because cell behavior is highly sensitive. So, considerations that lead to better harmonization of experimental conditions are timely and relevant, especially regarding the accumulation of large amounts of data in the literature [24].

Taherisadr and others presented in January 2013 another proposition, which is exhibited a mechanized red cells examination and grouping from created singular examples. This system based on shape highlights inner focal whiteness setup of red cells and their circularity with the assistance of decision logic. Red blood cells were classified into 12 classes to get results, diagnosis of blood issue, for example, iron inadequacy anaemia, the anaemia of chronic illness, β -thalassemia quality, sickle cell anaemia, haemoglobin C ailment, intravascular hemolysis, hereditary elliptocytosis, genetic spherocytosis and megaloblastic anaemia because of folic acid insufficiency can be conceivable. In this strategy, their portraying a few limits to characterization has high efficiency in the coordination of results. Really, they contemplated morphological perspectives and ascertained parameters and limits to get the quantitative consequences of cells regions and diameters [29].

Das and others in December 2012 presented a philosophy utilizing a few strategies of machine learning for describing RBCs in anaemia in view of microscopic images of peripheral blood smears. First, to reduce the unevenness of foundation brightening and noise, they preprocessed peripheral blood smear images taking into account geometric mean filter and the method of gray world assumption. At that point, watershed segmentation strategy applied to erythrocyte cells. Unhealthy RBCs, for example, sickle cells, echinocyte cells, teardrop cells, elliptocytosis cells, and solid cells classified based on their morphological shape changes. They observed that when a little subset of elements utilized by utilizing data pickup measures, the logistic regression classifier exhibited better in execution. They accomplished better forecast as far as general accuracy by 86.87%, sensitivity to 95.3%, and specificity was 94.13% [10].

In 2010, Hirimitugoda and Wijayarathna proposed a strategy to recognize thalassemia and malarial parasites in blood test images obtained from light magnifying instruments and explored the likelihood of quick and exact robotized finding of red blood cell issue. To assess the precision of the grouping in the acknowledgment of medicinal image patterns, they prepared two back proliferation Artificial Neural Network models (3 and 4 layers) together with image investigation systems on morphological components of RBCs. The three layers yielded the best execution with an error of $2.74545 e^{-005}$ and an 86.54% right accurate rate. The three prepared layers of an ANN comprised the most recent identification classifier to decide illnesses [17].

3 Overview on circular Hough transform

The Hough transform is classified as a popular feature extraction technique that transforms an image from its Cartesian to Polar coordinates. Duda and Hart in 1972 had used this technique of transformations after the presentation of Hough about it in

1962. The detection of circles/ellipses using Circular Hough Transforms (CHT) is more difficult than line detection algorithm due to a large number of parameters involved in describing the shapes. It is used to determine the parameters of a circle based on a known number of points that fall on the perimeter.

Three essential steps have done to be a CHT. One is Accumulator Array Computation, its cells constructed by those candidate pixels have high gradients and allowed to ‘votes’ process. Centre estimation is the second one that estimated by detecting the peaks in the accumulator array. The radius estimation step that using the same accumulator array [18, 19]. Figure 2a shows that each point in geometric space (solid dots) in the left generates a circle in parameter space (dashed circles) on the right. The circles in parameter space intersect at the (a, b) that is the centre of geometric space. In the same context, if the radius is unknown, then the points located in parameter space will fall on the surface of a cone. Each point (x, y) on the perimeter of a circle produces a cone surface in parameter space. The triplet (a, b, r) will correspond to the accumulation cell in the array where the largest number of cone surfaces intersect, as in Fig. 2b. A circle with a different radius has constructed at each level r .

In this paper, CHT has used to detect only the regular blood cells because they are near in shape to circle whereas sickle, burr and elliptocytosis have detected using their shape signatures.

4 Overview of neural networks and SVM

Neural Network (NN) is commonly known as a powerful function approximation for prediction and classification problems. Warren McCulloch and mathematician Walter Pitts who are first introduced a paper on how neurons might work in 1943 [20]. NN comprises input layer that contains input vector, hidden layers, and output layer with one neuron for every element of the output vector. Every element in the input layer is connected to every neuron in the hidden layers with weights associated with the connection between the input element and the hidden neuron, as shown in Fig. 3a. The back-propagation training process is to perform a particular function by adjusting the values of the connections (weights) between elements. The dataset divided into training and test subsets. The back-propagation works by presenting each input sample to the network where the estimated output be computed by performing weighted sums and transfer functions. It applies a weight correction to reduce the difference between the network outputs and the desired ones, in other words, the neural network can learn and can thus reduce the future errors [8, 17].

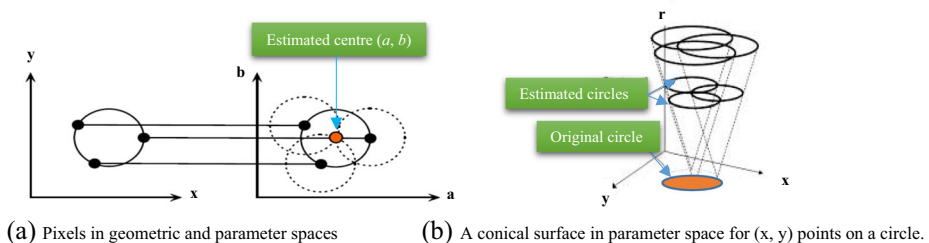


Fig. 2 An example of CHT

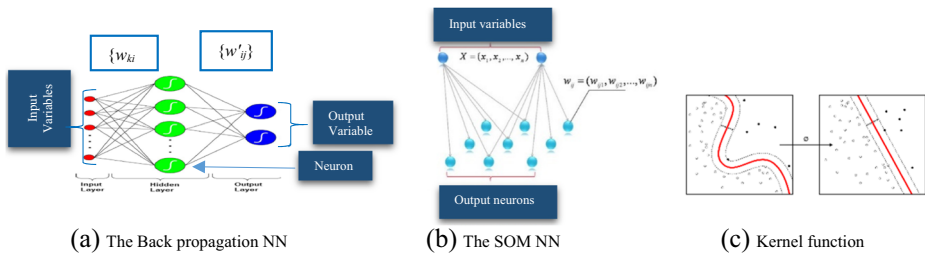


Fig. 3 The structure of the Back propagation, SOM neural networks and work of non-linear SVM

In the multilayer feed-forward network, an input vector is presented, and the output is compared with the target vector. If they differ, the weights of the network are changed slightly to reduce the error in the output. This is repeated many times and with many sets of vector pairs until the network gives the favourite output. However, training the Self-Organising Map (SOM) not requires any target vector. A SOM learns to classify the training data with no external supervision, as in Fig. 3b. The SOM is one of the most popular competitive learning neural network models. It developed by professor Kohonen [21]. He provided a way of representing multidimensional data in much lower dimensional spaces usually one or two dimensions. This process, of reducing the dimensionality of vectors, is essentially a data compression technique known as vector quantization. In addition, the Kohonen technique creates a network that stores information in such a way that any topological relationships within the training set are preserved.

The training process occurs in several steps and over many iterations. The weights of each node are initialized. A vector is presented to the lattice that randomly chosen from the set of training data. Every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU). After calculation the neighbourhood's radius of the BMU, this value starts large, typically set to the 'radius' of the lattice, but reduces each time-step. Any nodes found within this radius are considered being inside the BMU's neighbourhood. The weights of each neighbouring node (the nodes found in last step) are adjusted to make them more similar the input vector. The closer node is to the BMU. By repeating, the second step give the N iterations.

Support vector machine (SVM) constructs one or a set of hyperplanes in a multi-or infinite-dimensional space, which can be used for classification, regression, or other tasks. A hyperplane that has the largest distance to the nearest training-data point of any class is achieving good separation. Vladimir N. Vapnik and Alexey Ya. Chervonenkis were invented SVM algorithm in 1963. However, in 1992, Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes [6]. In this algorithm, every dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; although the classifier is a hyperplane in the transformed feature space, it may be nonlinear in the original input space. Figure 3c represents the work of kernel function of SVM.

In this paper, the traditional back propagation with one hidden layer, SOM networks and SVM have used to classify the same input vectors and comparing the results with the proposed algorithm which depends on shape signature method.

5 Cell shape signature

The method of shape signature has used in this paper through the detection process because it is simple in application on any closed geometrical shape and construct its own individual signature without depending on its size, position or dimensions [13]. It is summarized as; the distance between the boundary points of any shape and its centroid is the constructed shape own signature. After the operation of signature construction done, a process of matching among signatures begins in selecting shape that be described under the constraint rate of error.

For example, if the burr cell boundary points denoted by (x_{ijk}, y_{ijk}) , where i, j represent their components toward of X and Y-axis that related to the cell shape number k . The Euclidean distances D_{ijk} from its centroid C_k to (x_{ijk}, y_{ijk}) are calculated, plotted and saved as an individual's own signature. Since the normal blood cells have standard size and shape, the distance from centroids and boundary points are also equal, then they have to get the same own signature. Where C_{xk} , and C_{yk} are the centroid of cell k components on X and Y-axis. It is clearly, that the signatures of sickle, elliptocytosis and burr shapes differ completely from the normal blood cell signature because they have distinct shapes, as shown in Fig. 4.

Figure 4 shows that an example of signature related to normal, sickle, elliptocytosis and burr cells. In normal blood cells, all the distances from centroids to all the boundary points are equal (radii) that tend to the constructed signature is one straight line far from X-axis by the same measure of radii, as in Fig. 4a. In the proposed algorithm, the shape signature for normal cells is already done; but apply of CHT on them is faster, and then it has used to detect the normal cells. Other signatures for sickle, elliptocytosis and burr cell have presented in Fig. 4b. The step after the construction of cell signatures is saving in the system and matching with the input other signatures. In matching procedure, statistical equations related to the constructed signatures have applied on both cell signatures input and saved ones.

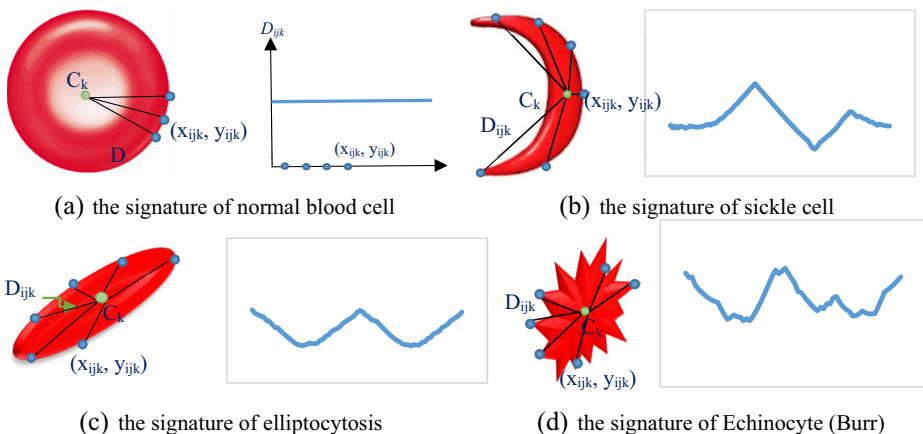


Fig. 4 The signatures for normal, sickle, elliptocytosis, and burr cells

There exist different cell shape size and orientation in any hue microscopic image of blood smear, then the signature data should be sorted and compute the absolute deviation, as in Eq. (1).

$$AVD = \frac{1}{n_{ij}} \sum_{i,j=1}^{n_{ij}} |x_{ijk} - \bar{x}_{ijk}| \tag{1}$$

For all k represent the number of cells in the image and i, j are the number of cells data boundary points. In Eq. (2), a method for measuring the least error by calculating the absolute value of subtraction between absolute deviations related to saved and input cell shape signatures [13].

$$DIF = ABS(AVD_{saved} - AVD_{input}) \tag{2}$$

The last decision of matching is based on the least error value of DIF, which gives the exact matched cell sickle, burr or elliptocytosis.

6 The proposed algorithm

The detection of deformed cells of anaemia disease is the main aim of proposed algorithm, beside the normal ones. It has concentrated on the famous and public three kinds of anaemia; sickle, elliptocytosis, and burr cells based on their shape signatures in a hued microscopic image of a human blood smear, as shown in Fig. 5.

Figure 5 illustrates the detected and contoured normal blood cells by using CHT with morphological functions. CHT has used to detect RBCs although they have on image borders or overlapped. At that point, the stuck and overlapped cells have separated by Watershed segmentation and the morphological functions. This process is done under constrains like the cell polarity; that it shows whether circulating blood cells are brighter or darker than the background. Computation method (Two-stage) has used to find out the accumulator array of CHT. It depends on figuring outspread histograms; radii are applied on the assessed cell focuses alongside the image [12]. Sensitivity factor is the sensor of the accumulator array in CHT. The recognition has included frail and halfway covered up or overlapped cells; higher estimations of the sensitivity expands the risk of false identification. The edge gradient

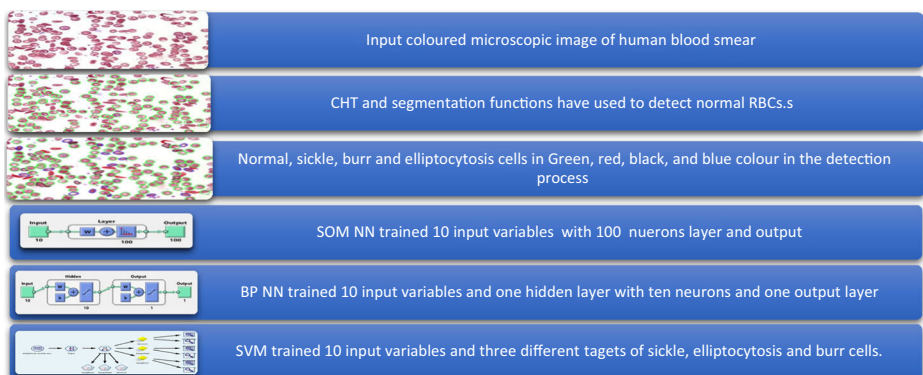


Fig. 5 The proposed algorithm in images

threshold; cells have largely a darker inside (cores) and encompassed by an outlying bright halo. The edge gradient threshold is extremely helpful for deciding edge pixels in these instances of image, both unhealthy and strong blood cells taking into account their difference have recognised well by setting a lower worth in the threshold. It distinguishes fewer cells of frail edges by expanding the estimation of the threshold [11].

Image segmentation is the process of selecting objects in the image from the background, i.e., partitioning the image into disjoint regions, such that each region is homogeneous regarding property, such as grey value or texture. The watershed transform is one of the famous segmentation methods. In more details if f is a digital grey value of an image. In the first, assuming that f is lower complete, each pixel which is not at a minimum has a neighbour of lower grey value. The lower slope $LS(p)$ of f at a pixel p , is defined as the maximal slope linking p to any of its neighbours of lower height, as in Eq. (3).

$$LS(p) = \frac{\max_{q \in N_G(p) \cup \{p\}} (f(p) - f(q))}{d(p, q)} \quad (3)$$

Where $N_G(p)$ is the set of neighbours of pixel p on the grid $G = (V, E)$ of V that is the set of all vertices and E is the set of all edges, whereas $d(p, q)$ is the distance associated to edge (p, q) (for $q = p$ the expression following the max operator is defined to be zero). With pixels whose neighbours are all of higher grey value, the lower slope is zero. The cost for walking from pixel p to a neighbouring pixel q is defined as:

$$cost(p, q) = \begin{cases} LS(p) \cdot d(p, q) & \text{if } f(p) > f(q) \\ LS(q) \cdot d(p, q) & \text{if } f(p) < f(q) \\ \frac{1}{2} (LS(p) + LS(q)) \cdot d(p, q) & \text{if } f(p) = f(q) \end{cases} \quad (4)$$

The distance $d(p, q)$, in Eq. (3) and (4), between non-neighbouring pixels p and q is defined as the minimum path length among all paths from p to q (this depends on the graph structure of the grid, in other words, the connectivity) [27]. In the proposed algorithm, the watershed segmentation has used to help in separation of all overlapped cells. Figure 6 shows that an example of using watershed in normal cells.

In Fig. 6a, the example of original image, the normal cells have filtered and selected in the grey image by using some morphological functions in (b). In Fig. 6c, the final image of watershed segmentation after divided the regions has appeared, then (d) shows the final image of contoured normal cells by green line after separating overlapped cells. Therefore, all separated cells are ready and available to get their measures like radii, areas, and so on. These measures have used in the application of neural networks later.

The pipelines of the proposed algorithm have illustrated in Fig. 7. In the first, the algorithm working on normal RBCs to enhance, select and contour them by using CHT with some of morphological functions and watershed segmentation. All their information data (radii, centres...) is ready now to create own variables. The filtration process has applied to remove all these normal cells from the input image and get the abnormal ones. The detection step of all abnormal cells (sickle, burr and elliptocytosis) has started by applying all previous steps again.

The data variables of them lead to construct their own signatures. These signatures have matched with the standard saved ones. Only three ideal shape standards have chosen from all samples for elliptocytosis, sickle, and burr cells under the same laboratory and microscope conditions, and all their measures have compared with other cells in images of whole samples.

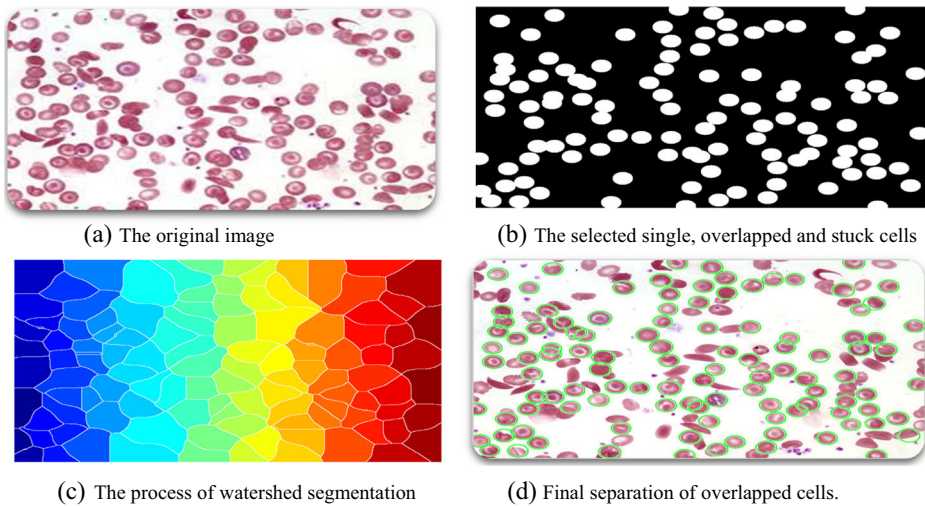


Fig. 6 The watershed and morphological functions application

All the measures concerned of these three standard cells have done to continue in the signature matching process. Table 1 shows all measures of these standard anaemia cells. The matching process has done by select the least resulted value from a comparison among all absolute deviation of them in the tested samples, as in Eq. (2). The BP NN, SOM NN and SVM have applied by using the input vectors of all the cells' extracted data.

The BP, SOM NNs and SVM have applied to make a comparison with the proposed algorithm and standing on the right stone for diagnosing and deciding whether the patient has one or more sorts of anaemia. The ten information data input variables are the Area, Convex Area, Perimeter, Eccentricity, Solidity, Ratio, AVD , DIF_{sis} , DIF_{bu} and DIF_{el} (see Eqs. 1, 2, 5, 6). Table 2. describes these variables in brief.

The proposed algorithm has stopped when it includes the input image covered by green, red, blue, black lines on the normal, sickle, elliptocytosis and burr cells, respectively. The previous input variables have been used for training BP NN, SOM NN and SVM. In SOM NN, no need to compute any targets for all input variables, however, BP NN and SVM should have targets to starting the train. Three output targets DIF_{sc} , DIF_e and DIF_{bur} have computed concerned of sickle, elliptocytosis, and burr cells, as in Eqs. (7), (8), (9), respectively.

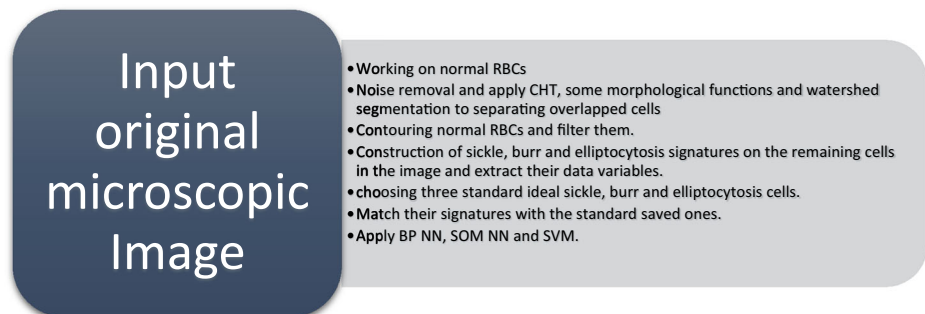


Fig. 7 Proposed algorithm steps

Table 1 The standard cells measures

Standard cells	Area	Convex Area	Perimeter	Eccentricity	Solidity	Ratio	AVD	DIF _{si}	DIF _{el}	DIF _{bu}
Elliptocytosis	460	485	86.76955	0.82451	0.948454	0.76777	2.211917	1.645007	0	0.06904
Sickle	234	462	102.669	0.728212	0.506493	0.278963	3.856924	0	1.645007	1.71405
Burr	350	395	80.42641	0.815844	0.886076	0.679956	2.142872	1.714052	0.06904	0

Table 2 Cells variables description

#	Variables	Description
1	Area	The actual number of pixels in the cell's region
2	Convex area	The number of pixels in convex hull of the cell, with all pixels within the hull filled in. In addition, the convex hull is $m \times 2$ matrix that specifies the smallest convex polygon, which can contain the cell region. Each row of the matrix contains the x-and y-coordinates of one vertex of the polygon.
3	Perimeter	The distance around the boundary of the cell's region. It has calculated the distance between each adjoining pair of pixels around the border of the cell.
4	Eccentricity	The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is in between 0 and 1. (0 and 1 are degenerate cases; an ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment).
5	Solidity	The relationship or proportion of the pixels in the convex hull that are also in the region, as in Eq. (5).
6	Ratio	It is a relationship between Area and Perimeter, as in Eq. (6).
7	AVD	It represents an absolute deviation of all values counted by the Euclidian distance from the cell centroid to its boundary points as in Eq. (1).
8	DIF _{si}	It represents the absolute difference among all AVDs related to input and saved standard sickle signatures, as in Eq. (2).
9	DIF _{bu}	It represents the absolute difference among all AVDs related to input and saved standard burr signatures, as in Eq. (2).
10	DIF _{el}	It represents the absolute difference among all AVDs related to input and saved standard elliptocytosis signatures, as in Eq. (2).

The variable Solidity (S_C) is the dividing the Area by its Convex Area for all separate cells as shown in the Eq. (5), whereas the variable that named Ratio (R_C) is calculated by dividing the Area of cells to Perimeter of cells, as shown in Eq. (6):

$$S_C = \frac{area_c}{(convex\ area_c)} \quad (5)$$

where $area_c$ is the area of each cell for all different cells, and $convex\ area_c$ represents the convex area of each cell for all detected abnormal cells.

$$R_C = \frac{4\pi(area_c)}{(perimeter_c)^2} \quad (6)$$

Here, $perimeter_c$ represents the perimeter of each cell for all kinds of detected abnormal cells. The target DIF_{SC} is an array that comprises binary numbers (1) and (0) on the values of DIF_{si} to classify sickle cell Anaemia, as shown in Eq. (7).

$$DIF_{SC} = \begin{cases} 1 & \text{if } DIF_{si} < 0.5 \\ 0 & \text{Otherwise} \end{cases} \quad (7)$$

In Eq. (7), if any of the absolute difference of absolute deviations values DIF_{si} is less than 0.5 then the decision is a sickle cell. Otherwise, when DIF_{SC} assumes zero value that is leading to ensure that cell is not sickle and may be one of other anaemia cells. The same way applies to target DIF_e , which is an array that comprises binary numbers (1) and (0) on values of DIF_{el} variable, as shown in Eq. (8).

$$DIF_e = \begin{cases} 1 & \text{if } DIF_{el} < 0.5 \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

When DIF_e is 1 the value of DIF_{el} is less than 0.5, which shows that the abnormal cell belongs to elliptocytosis. Otherwise, DIF_{el} value is greater than or equal 0.5 then DIF_e takes the value zero and leads to the decision that cell not classified as elliptocytosis and belongs to burr cells or other kinds of abnormal cells.

$$DIF_{bur} = \begin{cases} 1 & \text{if } DIF_{bu} < 0.5 \\ 0 & \text{Otherwise} \end{cases} \quad (9)$$

As the previous, DIF_{bur} is equal 1 when the variable value DIF_{bu} recorded value less than 0.5, which means the cell is surly burr. However, if DIF_{bu} has recorded value greater than or equal 0.5, the target has recorded zero and the network decided that the cell is not burr and may be belongs to other kinds of unchecked anaemia or other disease can also deforms RBCs.

The dataset has described as; 10 variable columns of inputs and 3 targets columns. The length of column is related to the size of cell in the tested image. The number of tested samples is 15; each one has 3 tested images, then the tested images have become 45 that have over 1000 cells at minimum. Table 3. illustrates the description of the formed abnormal cells information data for one example tested image. All variables have been ranged in the same types whereas only the output variables have valued binary values as 1 to exactly defined cells and 0 otherwise. The second column in Table 3. represents the intervals of variable measurements, for example, the Convex Area variable has a maximum value of one tested cell (1145) related to one image of one sample and minimum value (178) for the another cell in the same image. With knowing that, all the tested samples have the same distances from the microscope lenses (focal length), which give that all these measures have not big gaps from one sample to another.

The samples of the input variables have divided into 70% for training and 30% for testing, and the network has changed according to its error. The validation has used 15% out of samples to measure network generalisation and to pause training when generalisation stops improving. The remaining 15% out of all samples of cells have presented to be test samples that have no effect on training and so provides an independent measure of network performance during and after training. Ten variables trained and tested in the input layer, ten neurons of the hidden layer, and one neuron in the output layer. In addition, the mean square error (MSE) has applied which is defined as the average squared difference between outputs and targets, whereas lower values are better, zero means no error.

Table 3 Information data variables of cells

#	Variables	Type	Intervals of Measurements
1	Convex area	Ranged	178:1145
2	Perimeter	Ranged	51.1127:150.3675
3	Eccentricity	Ranged	0:0.97553
4	Areas	Ranged	158:869
5	Solidity	Ranged	0.506494:0.97914
6	Ratio	Ranged	0.278963:0.945956
7	AVD	Ranged	0.573935441:5.57356289
8	DIF_{si}	Ranged	$4.07924e^{-07}$:3.282988559
9	DIF_{bu}	Ranged	$5.26987e^{-11}$:3.430691048
10	DIF_{el}	Ranged	$3.31304e^{-10}$:3.361646293
11	DIF_{sc}	Binary	(1 for sickle cell, 0 for other normal/abnormal cells)
12	DIF_{bur}	Binary	(1 for bur cell, 0 for other normal/abnormal cells)
13	DIF_e	Binary	(1 for elliptocytosis, 0 for other normal/abnormal cells)

Every classification model has been assessed using three factual measures: classification accuracy, sensitivity, and specificity. These norms have set as true positive (TP), true negative (TN), false positive (FP) and false negative (FN). A true positive choice happens when the positive estimate of the classifier agreed with a positive expectation of the beforehand segmentation. A true negative choice happens when both the classifier and the segmentation proposes have the non-attendance of a positive expectation. False positive happens when the framework names sound cell (positive estimate) as an undesirable one.

False negative happens when the framework names a negative (abnormal) cell as positive. In addition, the classification accuracy has been characterised as a proportion of classified cells and equivalent to the entirety of TP and TN separated by the total number of RBCs (N) [5].

$$Accuracy = \frac{TP + TN}{N} \quad (10)$$

The sensitivity refers to the rate of arranged positive and is equivalent to TP partitioned by the total of TP and FN.

$$Sensitivity = \frac{TP}{TP + FN} \quad (11)$$

Specificity pointing to the rate of characterised negative and is equivalent to the proportion of TN to the entirety of TN and FP [5].

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

All tested images of samples are set far from the same distance (focal length) from the magnifying microscope optics to controlling all measures.

7 The experimental results

The flow of experimental results has done through three steps. One is the detection process for the normal and anaemia cells; the second is the signature matching to classify all these cells into sickle, burr, and elliptocytosis. The third step is applying of BP, SOM neural networks and SVM to make a comparison between them with the proposed algorithm. The morphological functions, watershed segmentation and CHT have used to enhance, separation of stuck cells, and noise removal from the image, then detect and contour the normal cells. The step after is filtering these detected normal cells from the image to select all stranger ones. The signature shape for each remaining cell in the resulted filtered image has constructed with its information data and saved. Therefore, the role of matching has come by test all signatures among themselves with standard ones to classify them as sickle, burr or elliptocytosis. To be on the right way, two kinds of neural networks have applied to all cells information data of the microscopic image to making other classification [5, 7].

All the blood cells samples have taken from patients that already have anaemia disease. The algorithm has applied on 45 chosen microscopic images with a size of 504×700 in 15 test samples with $40 \times$ magnifications to accurate detect normal, sickle, burr, and elliptocytosis cells. The cells have spread meagrely onto glass magnifying lens slides to permit odd kinds to be observed. Giemsa has used in blood preparations and stains red blood cells pink, platelets and white blood cells magenta. In the detection of normal cells, CHT has applied in the

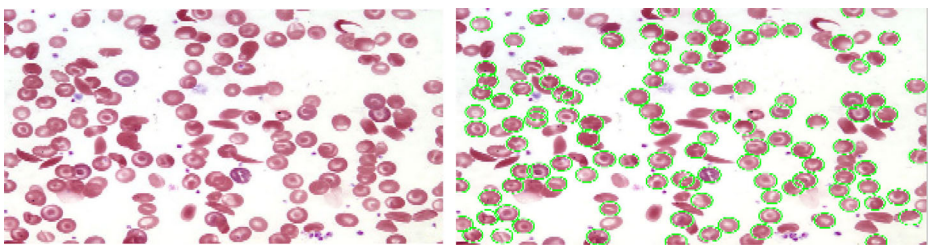
conditions of cell polarity to determine all dark and bright cells according to their intensity. Two-stage techniques are used after to compute the accumulator array of CHT; the sensitivity of this accumulator array in the proposed algorithm is 0.97 for brightness and 0.90 for dark cells, and an Edge gradient threshold 0.2 that is detected fewer cells with weak edges. These conditions help CHT to detect most of the normal RBCs that located singular, overlapped, and even that stuck to other abnormal cells. Figure 8 shows an example of input original image of human blood sample in (a), and the result of the contoured normal red blood cells.

In this input sample image, the total cells is 160 and only 109 normal cells have detected and contoured by green line, as in Fig. 8b. The remaining number of cells (51) may consider as detection errors, anaemia cells, platelets, white blood cells, or other kinds of diseases can attack RBCs. Platelets and white blood cells have been neglected (if they have appeared in the image) because the algorithm has only concentrated on normal and abnormal shapes of red blood cells [12, 14].

By removing all border cells from the images in Fig. 8a, only 153 RBCs have detected based on the proposed algorithm. The used filter for selection and remove normal cells comprises mask of balls simulate cells centroids and radii that measured between 17 and 20 pixels. Some difficulties have been happened through the flow of algorithm like normal/abnormal stuck cells, appearing platelets or blood impurity. The watershed process has applied to get the optimum cells separation and the morphological functions have worked to enhance the image and removing any platelets or impurities.

Now, it is the time to classify the remaining 44 cells into the three kinds of anaemia dependent on their shape signatures. Figure 9 presents the contoured abnormal cells except all of them on the image borders [14, 15]. The total number of detected anaemia cells is 42 out of 44 and it is classified as; 15 of sickle cells, 18 of elliptocytosis, and 9 of burr cells. The remaining number of those abnormal cells may be classified as other kinds of anaemia or cells stuck together to produce unknown shapes.

Figure 9 shows all detected elliptocytosis, sickle and burr cells have marked in clear serial numbers to be easier when they mentioned in the following tables. The 18 elliptocytosis have contoured by the blue line, the red line have represented the 15 sickle cells and all the burr cells have bounded by black line, as in Fig. 9 (a), (b), and (c). The construction of signatures that represents the detected elliptocytosis, sickle and burr cells have started. All these signatures have matched with the saved standard ones related to the previous Eqs. (1), (2). The following Tables 4, 5, and 6 illustrate the matching process among all signatures and measures of them. The constructed signatures of all the numbered elliptocytosis, sickle and burr cells in Fig. 9, the absolute deviation AVD in Eq. (1), and the least error DIF in Eq. (2) have counted. The decision of matching depends on the least error among the cell signatures and the standard



(a) Input original sample image

(b) Contoured normal cells only

Fig. 8 An example of the normal cell detection process

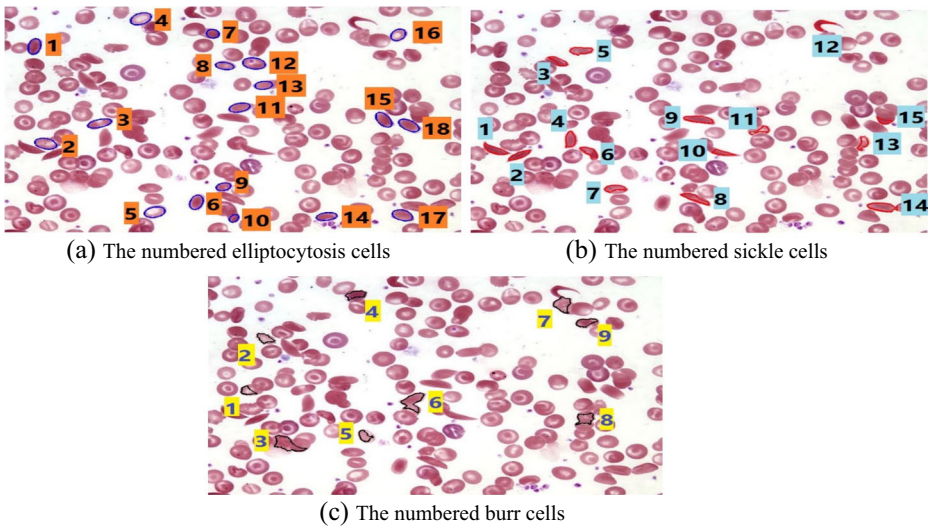


Fig. 9 Contoured and marked elliptocytosis, sickle, and burr cells in blue, red, and black lines respectively

chosen one. In fact, AVD and DIF have counted not for only these 33 detected cells but for all 44 cells to get the best match decision on them.

Table 4 describes the 18 elliptocytosis that appeared in the Fig. 9a. The cell number (11) has been chosen before as a standard signature cell of elliptocytosis. The values with red numbers in the row of *DIFel* are over 0.5, which is lead to out of matching with the standard one. Twelve cell shapes in Fig. 9a, which is represented one example image, have classified as elliptocytosis and the others may be belong to other kinds of anaemia, unknown cell shapes or not clear to be absolute elliptocytosis.


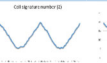

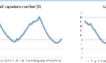
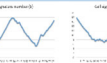
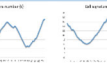
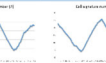

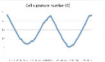
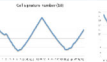
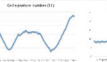


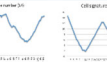

In Table 5, all the detected and numbered sickle cells in Fig. 9b signatures have drawn. The *AVD* and *DIF* have counted and given that, the cell number (12) has been considered as a standard signature for all images in whole samples, as mentioned before. Again, the row of *DIFel* has red values greater than 0.5, then the related signatures of these values not clearly sickle. Only six cells in image of Fig. 9b have classified as sickle cells and the others may be belong to other kinds of anaemia.

Table 6 shows all signatures and measures of selected and numbered burr cells in Fig. 9c. The data of *DIF* row give that only 5 cells out of 9 have classified as burr cell by the matching

Table 4 Elliptocytosis cells measures and signatures

Elliptocytosis cell signatures and measures									
Cell number	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Cell signature									
AVD	2.654921	2.04886	3.244861	1.764136	1.861919	2.116806	0.724981	2.156762	1.129325
DIFel	0.443004	0.163057	1.032944	0.44778	0.349998	0.095111	1.486935	0.055154	1.082591
Cell number	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]
Cell signature									
AVD	1.132543	2.211917	1.771723	1.69874	2.70218	2.65952	1.148013	1.915369	2.616559
DIFel	1.079374	0	0.440194	0.513177	0.490263	0.447603	1.063903	0.296547	0.404642

Table 5 Sickle cells measures and signatures

Sickle cell signatures and measures								
Cell number	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Cell signature								
AVD	4.272759	4.080093	3.517582	3.25253331	3.13415329	3.05439855	3.352836	5.573563
DIF _{si}	0.415835	0.223169	0.339342	0.60439	0.72277	0.802525	0.504088	1.716639
Cell number	[9]	[10]	[11]	[12]	[13]	[14]	[15]	
Cell signature								
AVD	4.863899	4.211205	3.478324	3.856924	3.119296	4.43175152	2.57680156	
DIF _{si}	1.006975	0.354282	0.378599	0	0.737628	0.574828	1.280122	


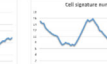




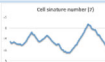
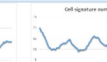

with the standard signature number (2). Obviously, the other values are belonging to signatures of shapes not clearly burr cell.

In Fig. 10, the detection process of normal cells in green lines and three of anaemia kinds; elliptocytosis in blue lines, sickle cells in red lines, and burr cells in black lines based on their signature shapes.

For about the neural networks application, SOM NN and BP NN have applied by using the same ten input variables. As it has been mentioned before, SOM NN has no need any target variable, however, BP NN needs to train these input variables three targets for three kinds of anaemia. These targets are DIF_e , DIF_{sc} , and DIF_{bur} for elliptocytosis, sickle, and burr cells, as in Eqs. (7), (8), and (9). Three BP NNs have applied with these targets individually. One is applied for elliptocytosis with ten input variables, ten neurons of one hidden layer, and one output layer. Figure 11 shows an example about first train of BP NN, which is represented elliptocytosis with an output target DIF_e . The training NN is shown in Fig. 11a, whereas the performance of best validation is 0.00070491 at epoch 31; the blue line represents training, the green one for validation, and the test with red line in Fig. 11b. The confusion matrices for training, validation, and testing processes of the first iteration BP NN have shown in Fig. 11c.

Table 7 presents a small part of results of elliptocytosis BP NN, which it contains the training iterations time, the number of iterations, the network performance, best validation performance, the number of epochs that the network has stopped, and the percentage of all confusion matrix. The confusion matrix has moved to swing in between two percentages along

Table 6 Burr cells measures and signatures

Burr cell signatures and measures						
Cell number	[1]	[2]	[3]	[4]	[5]	[6]
Cell signature						
AVD	1.855338	2.142872	4.804057	2.874954	2.042019	4.076434
DIF _{bu}	0.287534	0	2.661185	0.732082	0.100853	1.933563
Cell number	[7]	[8]	[9]			
Cell signature						
AVD	2.34371	2.002519	2.738587			
DIF _{bu}	0.200839	0.140353	0.595715			

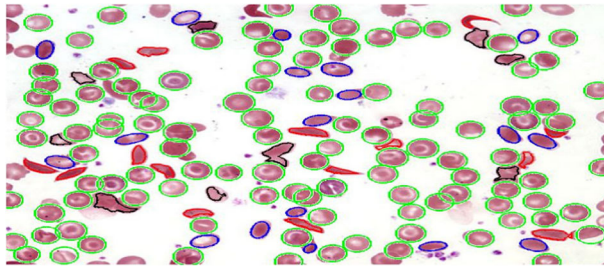


Fig. 10 Result of detected anaemia cells

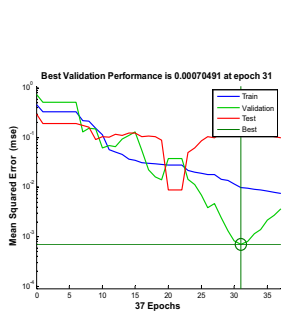
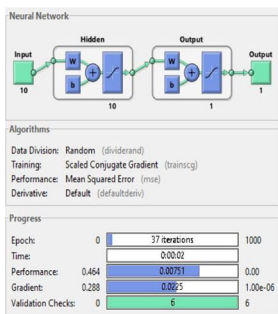
all the trains; 81% and 100%. In Table 7, the highest percentage has recorded 100% at zero seconds in the rows number 2 and 6; but the lowest one is 81.8% in the row number 4.

The second BP NN network has used the target DIF_{SC} of sickle cells data with the same input variables. Figure 12 presents another example about first training of sickle BP NN. The training NN has shown in Fig. 12a, whereas Fig. 12b, c have the performance of best validation and The confusion matrices for training, validation, and testing processes of the first iteration. The training has represented by blue line, the green one for validation, and the test with red line in Fig. 12b.

As in the previous, Table 8 introduces results of second BP NN that training the same input variables and different target. It contains the same columns of Table 7 with different values. The confusion matrix has recorded 100% as a highest value in the first train of the network after 32 iterations at zero seconds and repeated in rows number 2, 4 and 6; however, 88.6% has achieved as the lowest value in the column of confusion.

The target DIF_{bur} has been used in the third BP NN to train all input variables for classifying burr cells, as shown in Fig. 13. Figure 13a, b, c have given three images that represents an example of one train for the burr cells BP NN, the performance of best validation, and the confusion matrices for training, validation, and testing processes. Blue, green, and red lines have represented the training, validation, and the test.

Table 9 displays burr cells BP NN results, which has trained the same input variables and target DIF_{bur} . The same work as the last two tables has done in Table 9 with different trains values. The confusion matrix has recorded 100% as a highest value in all trains of the network after 52 iterations through zero seconds and repeated one time in row number 7; but the lowest recorded value 86.4% that has appeared in the fourth row.



(a) the training of BP NN process (b) the performance of best validation (c) training, validation, and testing confusion.

Fig. 11 An example of training elliptocytosis BP NN

Table 7 Results of first bp nn for elliptocytosis

Train	Time (seconds)	No. of iterations	Performance	Gradient	MSE of Training	MSE of Validation	MSE of Testing	Best validation	Epoch	Confusion %
1	2	37	0.00751	0.0225	$9.86162e^{-3}$	$7.04908e^{-4}$	$1.26711e^{-1}$	0.000705	31	97.7%
2	0	28	0.00150	0.00732	$1.24525e^{-2}$	$3.06499e^{-2}$	$8.41680e^{-3}$	0.03065	22	100%
3	9	6	0.0202	0.0479	$7.37702e^{-2}$	$1.00875e^{-1}$	$1.01276e^{-1}$	0.10088	3	93.2%
4	12	12	0.118	0.416	$1.74578e^{-1}$	$4.65893e^{-2}$	$4.98608e^{-2}$	0.046589	6	81.8%
5	11	13	0.0177	0.0431	$5.04140e^{-2}$	$7.41845e^{-3}$	$2.77016e^{-2}$	0.007418	7	93.2%
6	1.08	66	$8.70e^{-8}$	$8.87e^{-7}$	$8.69553e^{-8}$	$1.11850e^{-14}$	$6.29945e^{-3}$	$1.118e^{-14}$	66	100%
7	9	10	0.0274	0.0228	$1.36808e^{-1}$	$1.27459e^{-1}$	$1.32429e^{-1}$	0.12746	4	95.5%

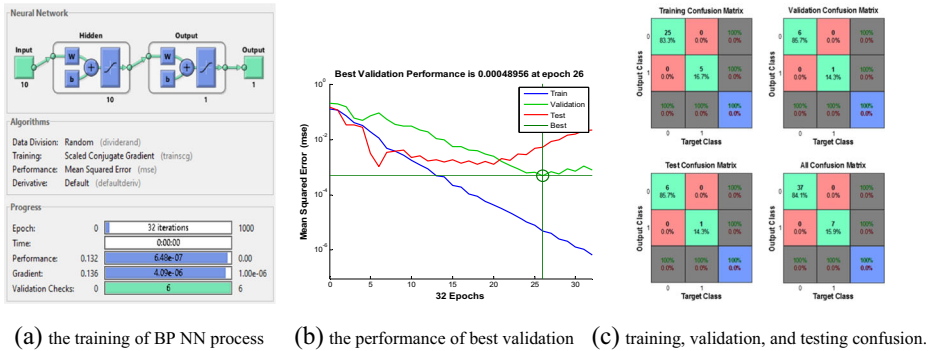


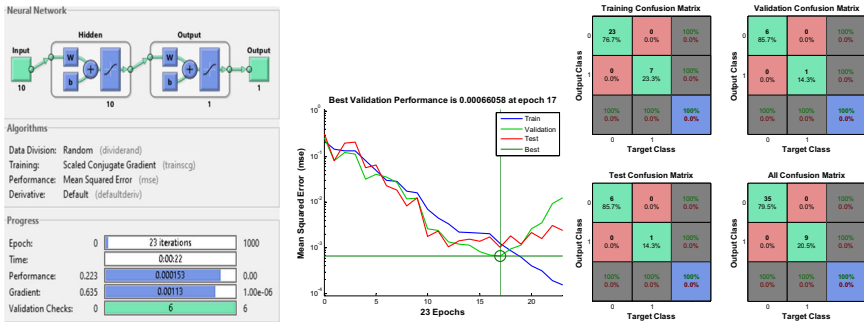
Fig. 12 An example of training sickle cells BP NN

One of the modern application in data classification is self-organising map (SOM). The topographic map has two important properties; at each stage of processing, each piece of incoming information is kept in its proper context/neighbourhood, and neurons dealing with closely related pieces of information are kept close together so they can interact via short synaptic connections. The interest is in building artificial topographic maps that learn through self-organisation in a neurobiological inspired manner. The spatial location of an output neuron in a topographic map corresponds to a particular domain or feature drawn from the input space. The SOM has a feed-forward structure with a single computational layer arranged in rows and columns. Each neuron is fully connected to all the source nodes in the input layer. As mentioned before, SOM NN has needed no target variables. By using the same variables in Table 3 except last three rows, which are targets for elliptocytosis, sickle, and burr cells, the SOM NN has trained to get comparison with BP NN on the proposed algorithm data. The training automatically has stopped when the full number of epochs have occurred. Figure 14 shows an example of training SOM NN has one layer, with neurons organised in a grid. When creating the network, the numbers the numbers of rows and columns in the grid have specified. Here, the number of rows and columns is set to 10. The total number of neurons is 100. The training runs for the maximum number of epochs, which is 200. For SOM training, the weight vector associated with each neuron moves to become the centre of a cluster of input vectors. Neurons adjacent to each other in the topology should also move close to each other in the input space, therefore it is possible to visualise a high-dimensional inputs space in the two dimensions of the network topology.

In Fig. 14b, the blue hexagons have represented the neurons. The neighbouring neurons have connected by red lines. The colours in the regions containing the red lines have showed the distances between neurons. The lighter colours have represented smaller distances, and the darker colours have represented larger ones. A band of dark segments has crossed from the lower-centre region to the upper-right region. The SOM network appears to have clustered the flowers into two distinct groups. The SOM weight positions in Fig. 14c shows the locations of the data points and the weight vectors after 200 iterations of the algorithm, the map has distributed through the input vector space. Figure 14d presents the neuron locations in the topology and shows how many of the training data are associated with each of the neurons (cluster

Table 8 Results of second bp nm for sickle cells

Train	Time (seconds)	No. of iterations	Performance	Gradient	MSE of Training	MSE of Validation	MSE of Testing	Best validation	Epoch	Confusion %
1	0	32	$6.48e^{-7}$	$4.09e^{-6}$	$4.89470e^{-6}$	$4.89559e^{-4}$	$5.38282e^{-3}$	0.022973	8	100%
2	14	14	0.00265	0.0118	$2.40703e^{-2}$	$2.29725e^{-2}$	$4.74877e^{-2}$	0.03065	22	100%
3	6	7	0.0374	0.0388	$9.93582e^{-2}$	$4.89269e^{-2}$	$2.82917e^{-1}$	0.048927	1	88.6%
4	17	19	0.000604	0.00266	$2.60857e^{-3}$	$6.17426e^{-2}$	$1.51918e^{-4}$	0.061743	13	100%
5	12	12	0.000802	0.00240	$9.03921e^{-3}$	$1.02125e^{-1}$	$1.09674e^{-1}$	0.10213	6	95.5%
6	24	26	0.000119	0.000641	$1.09208e^{-3}$	$2.45960e^{-2}$	$1.53982e^{-2}$	0.24596	20	100%
7	31	32	$7.63e^{-6}$	$6.23e^{-5}$	$6.74109e^{-5}$	$6.64959e^{-3}$	$1.43110e^{-1}$	0.00665	26	97.7%



(a) the training of BP NN process (b) the performance of best validation (c) training, validation, and testing confusion matrix.

Fig. 13 An example of training burr cells BP NN

centres). The topology is a 10 X 10 grid, so there are 100 neurons. The maximum number of hits associated with any neuron is 2. Thus, there are 2 input vectors in that cluster. In Fig. 14e, the neighbours typically have classified similar samples; neuron neighbour connections. The weight planes for each element of the input 10 vectors are in Fig. 14f. They are visualisations of the weights that connect each input to each of the neurons; such that; the larger weights have represented by darker colours. If the connection patterns of two inputs were very similar, you can assume that the inputs are highly correlated. In this example, input 1 has connections that differ greatly from those of input 4; but closer in 7 and 8.

SOM NN is easy to use and no need to target vectors but it has been slower in the trains followed the first one. For example, first train has taken 32 s; but the second train has taken 4 min and 47 s. The same input variables have used in all trains of SOM and BP neural network. All neural networks have agreed with the proposed algorithm in the classification process and the distinguishing of anaemia cells. In the detection process, the proposed algorithm has appeared effective when using cell signatures to select elleptocytosis, sickle, and burr cells and select them from the others, which are not clearly belonging to these three anaemia cells.

The support vector machine (SVM) has applied on the same input variables in Table 3. with the same three targets for elliptocytosis DIF_e , sickle DIF_{sc} , and burr cells DIF_{bur} , as in Eqs. (7), (8), and (9). All data of input variables divided into 70% for training and 30% for testing in a simple mode. Figure 15 shows the component nodes of SVM in a proposed stream. The stream is implemented in SPSS Clementine data mining workbench. Clementine uses client/server architecture to distribute requests for resource-intensive operations to powerful server software, resulting in faster performance on larger datasets. In Fig. 15a, *anaemia dataset node* is connected to EXCEL sheet file that contains the source data. *Type node* specifies the field metadata and properties important for modelling and other work in Clementine. These properties include specifying a usage type, setting options for handling missing values (the used dataset in this paper has not been missing values to handle), and setting the role of an attribute for modelling; input or output. *Type SVM node* enables to classify data into one of two groups without over fitting. The process has applied three times on three different targets to get good comparison and linear correlation between the outputs and the targets of Table 3. Figure 15b, c, d show the important variable in

Table 9 Results of third bp nm for burr cells

Train	Time (seconds)	No. of iterations	Performance	Gradient	MSE of Training	MSE of Validation	MSE of Testing	Best validation	Epoch	Confusion %
1	0	52	1.59×10^{-7}	6.42×10^{-7}	1.58514×10^{-7}	3.75585×10^{-8}	1.29586×10^{-3}	3.7558×10^{-8}	52	100%
2	34	34	8.55×10^{-5}	0.000642	4.99620×10^{-4}	5.72036×10^{-5}	1.26883×10^{-1}	5.7204×10^{-5}	28	97.7%
3	17	18	0.00501	0.0200	2.00624×10^{-4}	7.30672×10^{-4}	1.43650×10^{-1}	0.00073	12	95.5%
4	8	14	0.0601	0.123	8.85138×10^{-2}	2.75268×10^{-2}	2.11982×10^{-1}	0.02753	7	86.4%
5	15	17	0.00952	0.0265	3.56189×10^{-2}	1.41744×10^{-1}	1.45843×10^{-1}	0.14584	11	90.9%
6	13	18	0.000402	0.00183	7.76247×10^{-3}	6.29498×10^{-2}	4.48643×10^{-2}	0.06295	12	95.5%
7	22	23	0.000153	0.00113	1.24857×10^{-3}	6.60576×10^{-4}	1.03843×10^{-3}	0.000660	17	100%
8	28	49	0.000203	0.000460	6.31951×10^{-4}	9.56643×10^{-4}	5.74479×10^{-2}	0.000957	43	97.7%

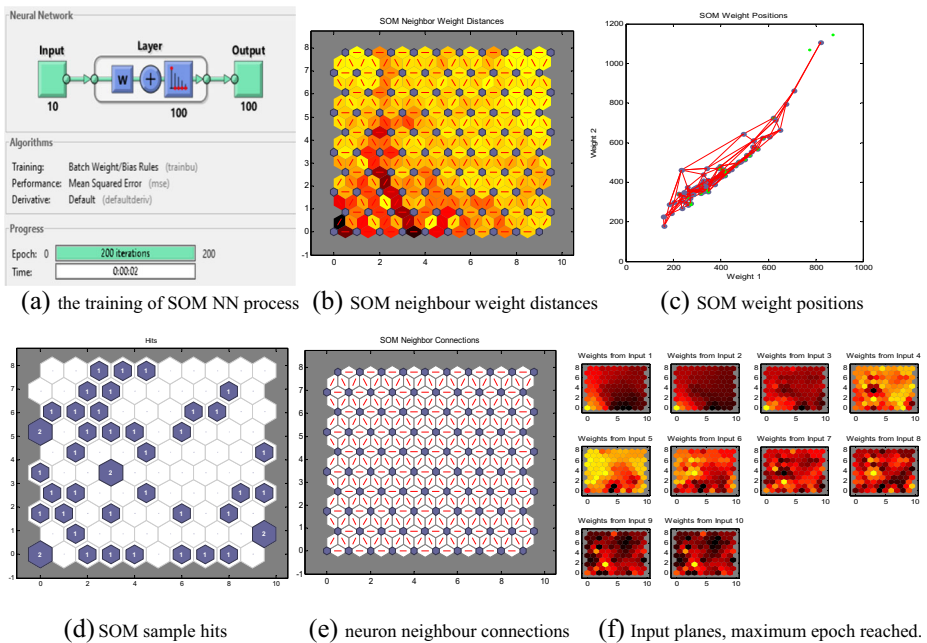


Fig. 14 An example of training SOM NN

the three cases of targets have been chosen. The variables DIF_ellipto and DIF_Burr have been chosen as the most important variables with the targets DIF_e and DIF_{burr} , respectively.

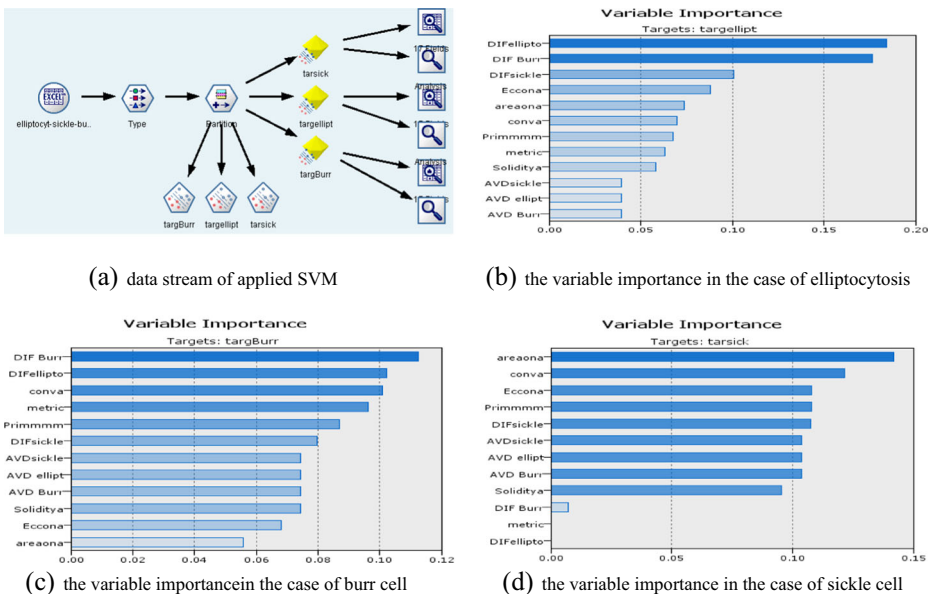


Fig. 15 The application of SVM and the variable importance

However, that the variable tends to the area of the cell has more important variable with the target DIF_{sc} . The linear correlations between the SVM outputs and DIF_e , DIF_{bur} and DIF_{sc} are 0.822, 0.631, and 0.799 for training and 0.916, 0.621 and 1.0 for testing, respectively.

The Matlab-2013a has used to build the algorithm on Windows 10 operating system with processor Intel ® Core™2Duo CPU T5550@ 1.83GHz and 2.50 GB RAM with 32-bit. The optical Nikon microscope has digitised all the images of blood samples.

8 Conclusions

In this approach, famous anaemia cells as elliptocytosis, sickle and burr have detected and contoured based on their own signatures. The cell signature has depended on its shape measures and created variables of their information. The Self-Organising Map (SOM), Back-Propagation (BP) neural networks and Support Vector Machine (SVM) have used these information data of cells' area, convex area, perimeter, eccentricity, solidity, ratio, absolute deviations, least error of elliptocytosis, least error of sickle, and least error of burr as input variables. In SOM no targets have needed; but in BP neural network and SVM three targets regarding to elliptocytosis, sickle and burr have used in training and testing. Some difficulties have faced the flow of the proposed algorithm, but by applying circular Hough transform, watershed segmentation and some of the morphological functions it has continued its way. The normal red blood cells also have detected their signatures measured, counted, and contoured in this approach. The performance has calculated by three statistical measures, classification accuracy, sensitivity, and specificity. Three BPs have been succeeding for about 100% as a highest value of all trains in confusion matrix about the three anaemia cells. The topology of SOM has 100 neurons, and the greatest number of hits associated with any neuron is recorded. In SVM, the highest value of linear correlation between its output values and algorithmic targets is 0.822 for training and 1.0 for testing. The proposed algorithm has tried to turn the light on an effective way for diagnosing and counting normal and anaemia cells in blood samples based on medical image processing.

References

1. Agroyannis B, Dalamangas A, Tzanatos H, Soubassi L, Fourtounas C, Kopelias I et al (1997) Alterations in echinocyte transformation and erythrocyte sedimentation rate during hemodialysis. *Artif Organs* 21:327–330
2. Aherne W (1957) The “burr” red cell and azotaemia. *J Clin Pathol* 10:252–257
3. Ahmed M, Patel AR (2015) Evaluation of normal reference range of schistocytes and burr cells in healthy adults. *Blood* 126:4540–4540
4. Anoop P, Parker-Williams J (2009) Morphological diagnosis of abetalipoproteinemia and the importance of a freshly prepared peripheral smear. *Eur J Haematol* 83:606–606

5. Bishop CM (2006) Pattern recognition and machine learning. Springer, Singapore
6. Boser BE, Guyon IM, and Vapnik VN (1992) A training algorithm for optimal margin classifiers, presented at the Proceedings of the fifth annual workshop on Computational learning theory, Pittsburgh
7. Chaudhuri B, Bhattacharya U (2000) Efficient training and improved performance of multilayer perceptron in pattern classification. *Neurocomputing* 34:11–27
8. Craven MW, Shavlik JW (1997) Understanding time-series networks: a case study in rule extraction. *Int J Neural Syst* 8:373–384
9. Da Costa L, Galimand J, Fenneteau O, Mohandas N (2013) Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders. *Blood Rev* 27:167–178
10. Das DK, Chakraborty C, Mitra B, Maiti AK, Ray AK (Feb 2013) Quantitative microscopy approach for shape-based erythrocytes characterization in anaemia. *J Microsc* 249:136–149
11. Deligiannidis L, Arabnia H (2014) Emerging trends in image processing, computer vision and pattern recognition vol. 1. Morgan Kaufmann, USA
12. Elsalamony HA (2014) Sickle anemia and distorted blood cells detection using hough transform based on neural network and decision tree. In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCVC), p. 1
13. Elsalamony HAM (2015) Comparing proposed signature with SURF in object detection process. *IETE J Res* 61:466–474
14. Elsalamony HA (2016) Detection of some anaemia types in human blood smears using neural networks. *Meas Sci Technol* 27:085401
15. Elsalamony HA (2016) Healthy and unhealthy red blood cell detection in human blood smears using neural networks. *Micron* 83:32–41
16. Freeman WJ (1991) The physiology of perception. *Sci Am* 264:78–85
17. Hirimutugoda Y, Wijayarathna G (2010) Image analysis system for detection of red cell disorders using artificial neural networks. *Sri Lanka Journal of Bio-Medical Informatics* 1:35–42
18. Illingworth J, Kittler J (1987) The adaptive Hough transform. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* pp 690–698
19. Illingworth J, Kittler J (1988) A survey of the Hough transform. *Computer Vision, Graphics, and Image Processing* 44:87–116
20. Karahoca A, Karahoca D, Şanver M (2012) Data mining applications in engineering and medicine, InTech, August
21. Kohonen T (1995) Learning vector quantization, in *Self-Organizing Maps*, ed: Springer, pp. 175–189
22. Lam P (ed) (2015) Anemia: causes, symptoms and treatments. *Medical News Today*. MediLexicon, Intl, UK
23. Lee H, Chen Y-PP (2014) Cell morphology based classification for red cells in blood smear images. *Pattern Recogn Lett* 49:155–161
24. Minetti G, Egée S, Mörsdorf D, Steffen P, Makhro A, Achilli C et al (2013) Red cell investigations: art and artefacts. *Blood Rev* 27:91–101
25. Mukherjee R (2014) Morphometric evaluation of preeclamptic placenta using light microscopic images. *Biomed Res Int* 9:293690
26. Mushabe MC, Dendere R, Douglas TS (2013) Automated detection of malaria in Giemsa-stained thin blood smears. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 3698–3701
27. Roerdink JB, Meijster A (2000) The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae* 41:187–228
28. Serebrennikova YM, Huffman DE, and Garcia-Rubio LH (2015) Characterization of red blood cells with multiwavelength transmission spectroscopy. *Biomed Res Int* p. 382641, 01/12
29. Taherisadr M, Nasirzonouzi M, Baradaran B, Mehdizade A, Shiraz I (2013) New approach to red blood cell classification using morphological image processing. *Shiraz E-Medical Journal* 14(1). <https://pdfs.semanticscholar.org/885e/661542fcb251eeea35c7484d54c9849a0187.pdf>. Accessed Jan 2013
30. Thirusittampalam K, Hossain MJ, Ghita O, Whelan PF (2013) A novel framework for cellular tracking and mitosis detection in dense phase contrast microscopy images. *Biomedical and Health Informatics, IEEE Journal of* 17:642–653
31. Tomari R, Zakaria WNW, Jamil MMA, Nor FM, Fuad NFN (2014) Computer aided system for red blood cell classification in blood smear image. *Procedia Computer Science* 42:206–213
32. Urrechaga E, Hoffmann JJ, Izquierdo S, Escanero JF (Jun 2015) Differential diagnosis of microcytic anemia: the role of microcytic and hypochromic erythrocytes. *Int J Lab Hematol* 37:334–340
33. Yi F, Moon I, Lee YH (2015) Three-dimensional counting of morphologically normal human red blood cells via digital holographic microscopy. *J Biomed Opt* 20:016005–016005



Hany A. Elsalamony was born in Cairo, Egypt. He received his B.Sc. (with honors) in mathematics, M.Sc. and Ph.D. degree in mathematics and computer science from Helwan University, Cairo, Egypt, in 2001 and 2006, respectively. He is currently a lecturer in the Department of Mathematics, Faculty of Science, Helwan University, since 2006. From 2007 to 2009, he was the Head of the Computer Science and Information Department, King Saud University, Saudi Arabia. Since 2007 to 2017, assistant professor of computer science in King Saud, Alkharj, Salman, Prince Sattam Bin Abdulaziz Universities, Saudi Arabia. He is a member in editorial board of many online and printed journals. Hany has collected different experiences in computer vision, data mining by working in many concerned projects. Hany published and reviewed several papers; all have focused on medical and digital image processing, and data mining, pure and discrete mathematics. His current research interests cover microscopic medical image processing, the data analysis and their applications on image understanding and signal processing.