Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN: 2521-3520 (online)
Volume 28, Issue 1, 2017
DOI: http: //doi.org/10.23851/mjs.v28i1.319

Research Article

Classification of Anemia Images Based on Principal Component Analysis (PCA)

Asma I. Hussein, Nidaa F. Hassan
Department of Computer Science, University of Technology, IRAQ.
*Corresspondant email: asma1139@yahoo.com or nidaaalalousi_5@yahoo.com

Article Info

Received
11/5/2016

Accepted
5/10/2016

Abstract

Blood cells are composed of erythrocytes (Red Blood Cells (RBCs)), the shape of RBC changes when the body suffers from different diseases such as Anemia. Classification of such diseases helps the medical technician to decide the type of Anemia in Laboratory analyzes in the hospitals. This paper proposed an automatic classification algorithm, which discriminates the different types of Anemia using Principal Component Analysis (PCA) algorithm and Decision tree. The proposed algorithm consists of four steps, at the first step preprocessing steps are applied on the RBC image, these RBC images then segmented in the second step, features are extracted using moment invariant in third step, this features are considered input to PCA so as to produced features vectors, at a final step features vector are inputted to Decision Tree to classify RBC image. Best classifications rates are (92%) obtained when using PCA algorithm compared with (74.1 %) which are obtained without applying PCA algorithm.

Keywords: Anemia, Principal Component Analysis (PCA), classification, Decision tree.

الخلاصة

خلايا الدم تتكون من خلايا الدم الحمر وشكل خلايا الدم الحمر يتغير عندما الجسم يعاني من مختلف الامراض مثل مرض فقر الدم. التصنيف في مثل هذا الامراض يساعد فني المختبر في تحديد نوع فقر الدم في المختبرات التحليلة في المستشفيات.في هذا البحث يقترح خوارزمية التصنيف الالي التي تميز انواع مختلفة من فقر الدم باستخدام تحليل المكونات الرئيسية و شجرة القرار، الخوارزمية المقترحة تتكون من اربع مراحل اول مرحلة هي معالجة الصورة التي تطبق على كريات الدم الحمراء و(RBC) تقسم في المرحلة الثانية، تستخلص الصفات باستخدام العزوم السبعة في الخطوة الثالثة وهذه الصفات تعتبر ادخال الى (PCA) التي تنتج متجهات من الصفات وفي المرحلة الاخيرة متجه الصقات تعتبر ادخال الى شجرة القرار لتصنيف صورة كريات الدم الحمر.

افضل نسبة تصنيف هي (92%) التي تم الحصور عليها باستخدام خوارزمية تحليل المركبات الرئيسية مقارنة مع نسبة (74%) التي تم الحصور عليها بدون تطبيق خوارزمية تحليل المركبات الرئيسية.

Introduction

One of the most classical main problems in image processing are Image classification based on attributes extracted from the input images, Type image can predict through images classification. The classification process depends on two steps: at the first step the classifier model is constructed to describe the predefined group of the image classes, the classifier model is created based on database by examining the tuples in the database defined by attributes. Each tuple is belonged to a class that is already defined, as determined by one of its attributes (class), at the second step, the constructed classifier model is used to classify the new entered images [1].

The Red Blood Cell (RBC) classification system can be used for educational purposes in medical schools and assist in the development of workers in the field of Hematology. Anemic can be classified according to RBCs shape distortion, size of RBC, color or hemoglobin content, depended on the types of Anemia suffered by the patient [2].

The current method used by the pathologists for identification of blood parameters is costly which are sometimes not affordable by the patient, also the generation of analysis report may require more time leading to loss of patient's life [3].

This paper proposed a new algorithm to classify Anemia using Principal Component Analysis

(PCA), (PCA) is one of the most popular appearance-based methods used mainly for dimensionality reduction in compression and recognition problems; since it reduces the size of training data which it entered to classifier [4]. The proposed algorithm is consisted of the following process: preprocessing, segmentation, feature extraction using Moment invariant, data reduction using Principal Component Analysis PCA and classifier using Decision Tree (ID3) to classify the types of Anemic.

Feature Extraction

Feature extraction is the process of defining a set of features, or image characteristics, which will most efficiently or meaningfully represent the information that is important for analysis and classification. Features are used to improve the effectiveness and efficiency of analysis and classification [5]. The main methods used to extract features in the proposed algorithm is moment's invariant, PCA is also used to reduce dimensions of image which will explain in the following sections, Decision tree one of the machines learning methods is used for classification decision.

Moment Invariants

Moments are considered the most widely used in the statistical theory, which display an averages or variations for distributions, its works on the description of image contents (or distributed) for the coordinates, thus it is designed to get geometric and detailed information about the image. Moments invariant are used in various image processing applications, such as: identify airplanes, matching of the scene, analysis of shapes, normalization of images, and discrimination of the letters written, determine the accuracy of the site and recuperation of images [6].

Hu in 1961 suggest the idea of using moments in shape recognition. The features of image remain fixed in the case underwent image to any changes (scale, translation and rotation) when use a set of moment invariants, the following steps illustrate the computation of moment invariant:

Step1: Compute the central moments, it define: the moments that have the property of translation invariance and are denoted byμ_{pq}, it is defined in Equation 1:

$$\mu_{pq} = \sum_{x=0}^{x=N-1} \sum_{y=0}^{y=M-1} (x-\bar{x})^p . (y-\bar{y})^q f(x,y) \tag{1}$$

Where $\bar{x}$ and $\bar{y}$ are the coordinates of the centered, and they are calculated using Equation (2) and Equation (3).

$$\bar{x} = \frac{m_{10}}{m_{00}} \tag{2}$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \tag{3}$$

The following Equations are computed to easily check that the central moments reach to up the order p+q ≤ 3, Equation (4) to Equation (13) [6]:

$$\mu_{00} = m_{00} \tag{4}$$
$$\mu_{10} = 0 \tag{5}$$
$$\mu_{01} = 0 \tag{6}$$
$$\mu_{20} = m_{20} - \bar{x}m_{10} \tag{7}$$
$$\mu_{02} = m_{02} - \bar{y}m_{01} \tag{8}$$
$$\mu_{11} = m_{11} - \bar{y}m_{10} \tag{9}$$
$$\mu_{30} = m_{30} - 3\,\bar{x}m_{20} + 2\,\bar{x}^2 m_{10} \tag{10}$$
$$\mu_{12} = m_{12} - 2\,\bar{y}m_{11} - \bar{x}m_{02} + 2\,\bar{y}^2 m_{10} \tag{11}$$
$$\mu_{21} = m_{21} - 2\,\bar{x}m_{11} - \bar{y}m_{20} + 2\,\bar{y}^2 m_{01} \tag{12}$$
$$\mu_{03} = m_{03} - 3\,\bar{y}m_{02} + 2\,\bar{y}^2 m_{01} \tag{13}$$

Step2: normalized central moments using to get the Scale invariance η_{pq}, as Equation (14) & Equation (15):

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}{}^{\gamma}} \tag{14}$$

Where:

$$\gamma = \left[\frac{(p+q)}{2}\right] + 1 \tag{15}$$

Step3: From normalizing central moments for order three, that are invariant in the case underwent image any changes (scale, translation and rotation) compute seven moments are given as Equation (16) to Equation (22) [7]:

$$\emptyset_1 = \eta_{20} + \eta_{02} \tag{16}$$
$$\emptyset_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{17}$$
$$\emptyset_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{18}$$
$$\emptyset_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{19}$$

Volume 28, Issue 1, 2017 DOI: http: //doi.org/10.23851/mjs.v28i1.319

$$\emptyset_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (20)$$

$$\emptyset_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (21)$$

$$\emptyset_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (22)$$

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is the most famous methods that have been used in image classification. PCA can remove noise and redundant information by computing the most important elements and structure of the original data. It can be used to compress the data size, by reducing the dimensions of image, without much loss of information [8].

PCA used to identify the image features and competences for improving the accuracy of Red Blood Cell image analysis through reduce the number scopes of image.

The Proposed Algorithm

In this proposed algorithm, Anemia in digital images is recognized based on Moment Invariant and PCA. The classifier ID3 is used to classify types of Anemia. The structure of the proposed algorithm is composed of two phases (Training phase and testing phase) as shown in Figure 1; each phase has specific tasks which illustrated in the following sections.

Pre-processing

This stage consists of the following three steps:

Collection of Anemia Dataset:

Due to the lack of available standard Anemia images, Anemia Dataset is collected from number of hospitals: *Al-Karama Teaching* Hospital, *Al-Yarmouk Teaching* Hospital, and *Ibn Al-Balady* Hospital. Images are captured in the JPEG format with the maximum resolution.

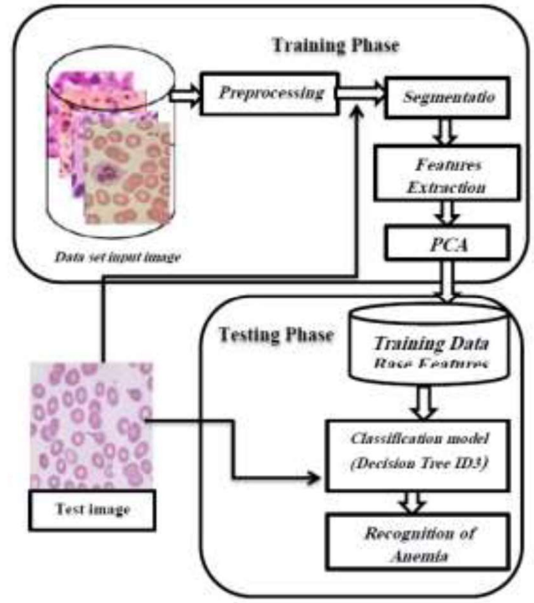


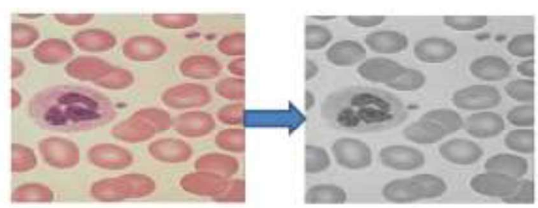
Figure 1: Block diagram of Proposed Algorithm.

Image Conversion:

The color (RGB) images are converted to the gray scale as shown in Figure 2.

Noise Removal:

Removal of noise is done by using median filter so as to improve image quality, while preserving the integrity of edges and detail information.



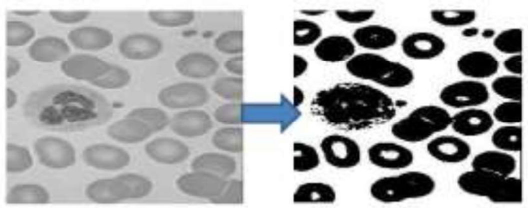
(a) Original image (b) Gray image

Figure 2: Conversion of color image to gray image.

Image Segmentation

In this stage, subdivides of an (RBC) Red Blood Cell image is achieved to produced number of regions, each region represent one cell from smear of blood, Otsu's thresholding method is used to segmented image by divided the pixels into two clusters (object and background) based on threshold. Histogram based on clustering is used to find the threshold value; the following steps are illustrated this process:

1. Compute the total of histogram.
2. Compute sum histogram for each background and object.
3. Find the weight and mean for each class.
4. Compute the Variance between two classes.
5. Finally the threshold value considers maximum value of variance between two classes. Figure 3 shows the segmented cells.



(a) Entire image. (b) Isolated cells.
Figure 3: Segmentation of cells in RBC image.

Extracted Features of Anemia Images

In this stage the recognition of segmented objects or regions from other objects is accomplished, so each object consider as pattern with its features, the features are extracted to reduce the original data set and to distinguish one input pattern from another.

Moment invariant is used to extract features from images, these features are important for image classification process; Table 1 as shown below; the numerical value features of Anemia images.

Principle Component Analysis (PCA) of Anemia Images

PCA is used to decrease the dimensional representation of Anemia image, after extract seven features from moment invariant, these features are inserted as input to PCA and then PCA produced Eigen value and eigenvector. In classification process, each training image is transformed into a vector. The covariance matrix is computed by multiply variance image by variance image. Eigenvectors associated with eigenvalues are constructed; represent the features vector of Anemia image which as input to Decision tree (ID3) algorithm to classification images of Anemia.

Table 1: Show the numerical values of feature values and classes.

	Image name	$\emptyset_1$	$\emptyset_2$	$\emptyset_3$	$\emptyset_4$	$\emptyset_5$	$\emptyset_6$	$\emptyset_7$	Class
1	normal	0.24	0.552	0.891	1.263	2.951	4.31	0.7	normal
2	Iron	0.246	0.481	0.888	1.946	3.286	1.09	5.3	Iron
3	Sickle cell	0.268	0.551	0.902	0.91	2.366	1.18	6.8	Sickle cell
4	Target cell	0.37	0.617	0.942	0.769	1.984	1.24	6.4	Target cell
5	normal	0.293	0.97	0.935	0.587	0.871	1.16	6	normal
6	Hypochromic	0.297	0.613	0.896	1.121	2.635	1.11	5.9	Hypochromic
7	Teardrop	0.308	0.869	0.937	0.609	0.974	1.08	6.2	Teardrop
8	normal	0.336	0.359	0.885	1.689	3.53	1.20	6.9	normal
9	Target cell	0.362	0.722	0.923	0.872	1.794	1.04	6.9	Target cell
10	Ovalocytes	0.394	0.642	0.909	0.751	1.936	1.14	6.7	Ovalocytes
11	Teardrop	0.6	0.395	0.888	2.258	3.512	1.10	7	Teardrop
12	Iron	0.728	0.425	0.778	1.067	3.117	1.25	6.6	Iron

Decision Tree and Decision Rules

After feature extraction in the previous stage, classification consider final stage where perform building a Decision Tree and Decision Rules depended on features in the training phase for Anemia images, to classify new image in the testing phase pass to all stage in the training phase to given decision type of *Anemia*.

The categorical values of the seven features using moment invariant which classify by the ID3 algorithm are described in Table 2. This Table consists of three fields: *Attributes*, *Attribute-value*, and *Range of Values*.

Table 3 illustrates different samples of categorical values of features using moment invariant.

Table 2: Categorical values of features using Moment invariant.

Attributes	Attribute value	Range of Values
Ø1(A)	A1	[0.12 To 0.402]

	A2	[0.403 To 0.684]
	A3	[0.685 To 0.966]
	B1	[0.711 To 0.804]
Ø2(B)	B2	[0.805 To 0.897]
	B3	[0.898 To 0. 99]
	C1	[0.302 To 0.528]
Ø3(C)	C2	[0.529 To 0.754]
	C3	[0.755 To 0.98]
	D1	[0.176 To 1.412]
Ø4(D)	D2	[1.413 To 2.648]
	D3	[2.649 To 3.884]
	E1	[0.439 To 1.045]
Ø5(E)	E2	[1.046 To 1.651]
	E3	[1.652 To 2.257]
	F1	[4.029 To 4.418]
Ø6(F)	F2	[4.419 To 4.808]
	F3	[4.809 To 5.109]
	G1	[0.673 To 0.707]
Ø7(G)	G2	[0.708 To 0.741]
	G3	[0.742 To 0.775]

Table 3: Sample of Categorical values of features using Moment invariant.

	$Ø_1$	$Ø_2$	$Ø_3$	$Ø_4$	$Ø_5$	$Ø_6$	$Ø_7$	Class
1	A1	B3	C1	D3	E2	F2	G6	normal
2	A1	B2	C1	D3	E3	F4	G4	Iron
3	A1	B3	C2	D3	E1	F2	G6	Sickle cell
4	A1	B3	C2	D2	E1	F4	G5	Target cell
5	A3	B2	C1	D3	E3	F6	G3	normal
6	A1	B3	C1	D2	E1	F6	G3	Hypoch romic
7	A1	B3	C3	D1	E1	F5	G4	Teardro p
8	A1	B2	C1	D3	E2	F3	G6	normal
9	A1	B3	C3	D2	E1	F3	G5	Target cell
10	A2	B2	C3	D2	E1	F5	G3	Ovaloc ytes
11	A1	B3	C3	D1	E1	F5	G2	Teardro p
12	A3	B1	C1	D3	E1	F4	G4	Iron

After the decision tree is constructed based on select the best attribute to branch using Entropy and Information Gain, it is easy to form the set of rules as follows:

R1: IF (Q1=A1) Then Class=Normal.
R2: IF (Q6=F4) and (Q2=B3) Then Class=Target cell.
R3: IF (Q4=D1) and (Q1=A1) and (Q2=B3) Then Class=Tear drop.
R4: IF (Q2=B3) and (Q7=G3) and (Q4=D2) and (Q3=C1) Then Class=Hypochromic.
R5: IF (Q7=G6) and (Q2=B3) and (Q5=E2) and (Q6=F2) Then Class=Iron.
R6: IF (Q1=A2) and (Q3=C3) and (Q5= E1) and (Q7=G3) Then Class= Ovalocytes.

Results and Discussion

Experimental evaluation of the proposed algorithm is performed on two stages (training and testing).
The training images in this algorithm consists of 7 classes (Teardrop, Target cell, Ovalocytes, Sickle cell, Normal, hypochromic and Iron deficiency), each class consist of different number of images. Total number of images is (455), where (280) images are selected for training phase and (175) images are selected for testing phase. Accuracy measure is the most widely used measurement method to evaluate the classification images, thus to compute the accuracy of the decision tree the following Equation is used:

$$Accuracy = \frac{\textbf{Number of Correctly Classified Images}}{\textbf{Total Number of Images}} \times 100\%$$

Table 4 shows the accuracy rate of features extracted before using PCA, and Table 5 shows the accuracy rate of features extracted by using PCA. The optimal results of recognition are achieved using PCA with rate (92%) compared with rate (74%) when classifier before apply PCA.

Table 4: Accuracy rate before PCA.

Type of Classification	Correct	Wrong	Accuracy
Teardrop	80	13	71%
Target cell	8	2	80%
Ovalocytes	30	3	80%
Sickle cell	6	4	60%
Normal	45	11	80%
Hypochromic	80	10	87%
Iron	90	57	61%

deficiency	
Average	74.1%

Table 5: Accuracy rate based on PCA.

Type of Classification	Correct	Wrong	Accuracy
Teardrop	85	5	94%
Target cell	12	3	90%
Ovalocytes	27	3	90%
Sickle cell	14	1	93%
Normal	55	1	98%
Hypochromic	85	5	94%
Iron deficiency	125	22	85%
Average			92%

Conclusion

In this paper, classification of Anemia images algorithm is presented, this algorithm includes applying four steps: preprocessing RBC images, then RBC images are segmented into two cluster (object and background), feature are extracted from segmented images using Moment invariant, also PCA is used to reduce image data, at last the features are classified into classes using Decision Tree (ID3) algorithm. Accuracy results indicate that: (92%) accuracy rate is obtained when PCA is applied compared with (74%) is obtained without applying PCA.

References

[1] H. Mazien, "Proposed system for the diagnosis of skin diseases using Multiwavelet Transform and Decision Tree," M.Sc. Thesis, University of Technology, Department of Computer Science, 2015.

[2] Jameela Ali, *et al.*, "Red Blood Cell Recognition using Geometrical Features," *International Journal of Computer Science Issues,* vol. 10, no. 1, 2013.

[3] S. Chandrasiri, *et al.*, "Morphology Based Automatic Disease Analysis Through Evaluation of Red Blood Cells," in *Department of Information Technology/ Sri Lanka Institute of Information Technology, 2014 Fifth International Conference on Intelligent Systems, Modelling and Simulation*, 2014.

[4] M. Hussein, *et al.*, " Image Authentication Using PCA And BP Neural Network," *Eng.& Tech. Journal,* vol. 28, no. 22, 2010.

[5] Z. F., "Mammogram Images Classification Using Contourlet Transform," M.Sc. Thesis, University of Technology, Department of Computer Science, 2014.

[6] K. Hosny, "Exact and fast computation of geometric moments for gray level images," *Applied Mathematics and Computation Journal,* vol. 189, p. 1214–1222, 2007.

[7] D. N. George, "Tumor Type Recognition Using Artificial Neural Networks," M.Sc. thesis, Iraqi Commission for Computers And Informatics Institute for Postgraduate Studies, Baghdad, 2013.

[8] R. M, "Classification of Face Image Based on Gender Using Intelligent Method," M.Sc. Thesis, University of Technology, Department of Computer Science, Baghdad, 2015.