RESEARCH ARTICLE OPEN ACCESS

An Intelligent System for Diagnosing Anemia Diseases

R.A.E. Ibrahim [1], A. E. E. El Alfi [2], A.A.E. Abdallah[3]

[1],[2],[3] Computer Science Department, Mansoura University/ Faculty of Specific Education - Egypt.

ABSTRACT

This paper presented an intelligent system for diagnosing anemia diseases such as Sickle Cell Anemia, Iron Deficiency Anemia, Vitamin and Folic Acid Deficiency Anemia, Aplastic Anemia, Anemia caused by Bone Marrow Disease and Hemolysis Anemia. The proposed system uses color coherence vector (CCV) method: for extracting the features of images acquired using microscope connected to a digital camera, scientific references for blood analysis such as (Clinical Hematology Atlas) in its various versions, and search engines such as (Yahoo, Google, Bing, etc.). The results showed a significant similarity between the laboratory analysis and the diagnosis of the proposed system. Which indicates the accuracy and high efficiency of the proposed system.

Keywords: Blood cell, Anemia, Image processing, Color Coherence Vector, intelligent system.

I. INTRODUCTION

Blood is a specialized bodily fluid that delivers necessary substances to the body cells, such as nutrients and oxygen and transport waste products away from these same cells. [1]

Anemia is defined as a decrease in erythrocyte count or hemoglobin (Hb) concentration.[2]The types of anemia are: Microcytic Anemia:(Sickle cell anemia, Iron Deficiency Anemia)[3], Macrocytic Anemia(Folic Acid Deficiency), Normocytic Anemia(Aplastic, Anemia of Chronic Disease, Hemolytic Anemia).[4] Intelligent systems help users to make decisions, as smart decision is the key technology of smart systems.[5] One of the commonly known techniques used in image retrieval is color coherence vector (CCV) which is an image descriptor (or more specifically, a color descriptor) that extracts color-related features from an image [6,19].

Therefore, this paper presents a system to identify one of the types of blood diseases, anemia, using image retrieval.

This paper is divided as follows:

Part one: presents a general summary of anemia, Part two: proposed system is described, Part three describes a discussion of the experimental results. Finally, the conclusion and references.

II. RELATED WORK

S Y Veronica, et al. have suggested an information method based on a web rule. using the rule-based algorithm to facilitate monitoring and detection of anemia among expectant mothers exposed to extreme risks. Obtained results confirming the effectiveness of the system is 90%. [7]

M. VadivelV, et al. have discussed the automatic classification of SCA system (Sickle cell anemia). by, the images entered to the candidate gaussian, then energy feature was used to extract the features, and KNN method is used for classification. The results of the system showed an accuracy of up to 93%.[8]

Bhavinkumar A. Patel and jay Parikh have presented a model that helps predict anemia in the initial stages using the Support Vector Machines (SVM). The results proved that the model is effective compared to other methods.[9]

Shilpa A. Sanap, Meghana Nagori and Vivek Kshirsagar They presented a classification analysis and prediction of anemia in patients using data mining techniques, it worked with C4.5 classification method, decision tree algorithm, support vector machine implemented as J48, and minimum sequential optimization (SMO) in Weka. Note that the C4.5 algorithm has better performance and higher accuracy.[10]

Laith Alzubaidi, et al. have introduced neural network to classify Red Blood Cells (RBCs). In (3) categories: abnormal ('S'), normal ('N') and miscellaneous ('M'). then the feature extractor then applied an error-correcting output codes (ECOC) for the classification mission. The results showed high accuracy up to 92.06%.[11]

Serhat kilicarslan, mete celik and safak sahin have proposed two hybrid models using deep learning algorithms and Genetic Algorithm (GA)of Stacked Auto Encoder (SAE) and Convolutional Neural Network (CNN) to predict HGB anemia and nutritional anemia (iron deficiency anemia, B12 deficiency anemia and folic acid), the performance of GA-CNN algorithm, by a 98.50% accuracy.[12]

Ning Zhang, et al. have proposed an effective method based on dominant colors (DCs) and color moments (CMs) for image retrieval. different similarity measure methods were carried out based on the DC feature and CM feature. The average precision and recall were up to 87% and 44%, respectively. [13]

Mahsa Moslehi and Felipe P. J. de Barros have Suggested an approach based on color coherence vectors (CCV) The sensitivity of the CCV to spatial information makes it a suitable metric for evaluating the performance of data ingestion technologies. Under various factors, such as the number of measurements and structural parameters of the log conductivity field, they compare the performance of CCV with root-mean-square error (RMSE).[14]

Thoopsamut, et al. have developed a method to verify user identity by using handwritten signature. Color Coherence Vector (CCV) and signing behavior parameters with the stylus pen on android smartphones are used. The experimental results showed accuracy level of 94%. [15]

Yibo Li and Mingjun Liu have fusion classification algorithm employed rotation uniform invariant LBP and color coherence vector was proposed for optical aerial images. Support Vector Machine (SVM) based on RBF was applied to classify image. The experimental results 96.66%. [16]

S. B. Mallikarjuna, et al. have proposed a new method for classification of arecanut images of different diseases. They have classificated four categories, namely, images of Healthy, Rot, Split and Rot-Split for classification. They discovered the combination of multi-gradient and AlexNet by feeding enhanced images as input. The results of the Implementation of the proposed approach are superior.[17]

III. ANEMIA

Anemia or lack of red blood cells/hemoglobin in the blood that leads to pallor and fatigue is one of the main challenges in Egypt. It affects 27.2% of children under five and 25% of women of reproductive age (15-49 years). Anemia during pregnancy is one of the main causes of anemia in infants and children.[18]

The general health of the human body is evaluated through the report CBC. It is used in the diagnosis of many diseases, most notably anemia. [9]

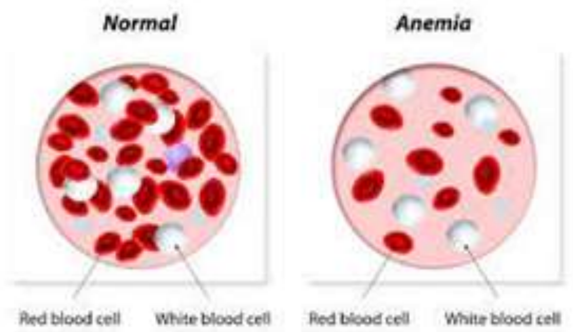


Fig 1: The difference between a normal person's blood sample and an anemic sample

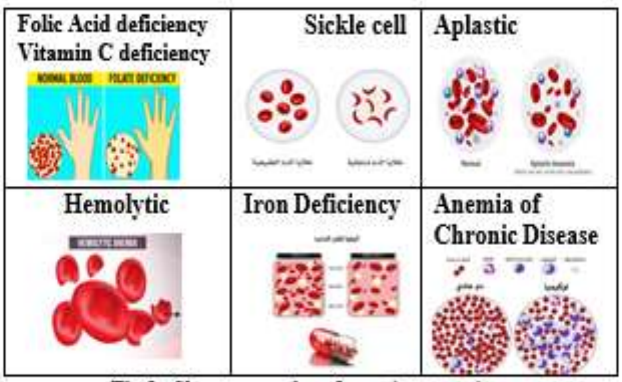


Fig 2: Shows examples of anemia categories

IV. PROPOSED APPROACH

The proposed system steps will be explained in the following section:

The system consists of four parts:

1- Building Knowledge base with images of the types of anemia previously described.

2- Using a treatment process method (CCV) to extract features.

3- Performing the matching process to extract the similar images.

4- Determining the appropriate type of anemia (Diagnosis).

The system parts will be explained in detail in the following part:

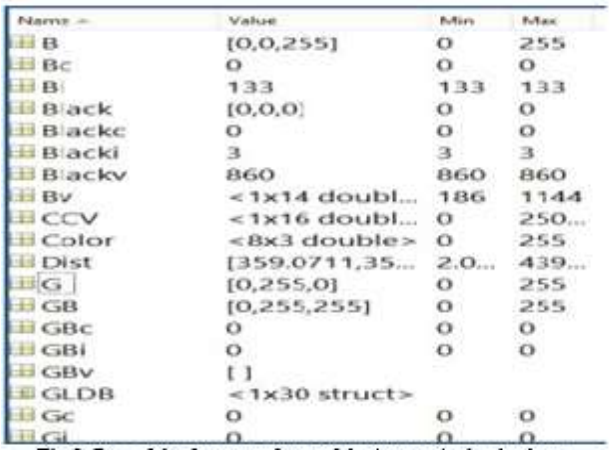
First, building the Knowledge base:

The knowledge base is drawn from human experts and laboratory labs.

Table1. Some of the main types of anemia and their sub-types.

The main types	Categories
Microcytic Anemia	Sickle Cell Anemia. Iron Deficiency Anemia .
Macrocytic Anemia	Folic Acid Deficiency. Vitamin C Deficiency.
Normacytic Anemia	Aplastic - Anemia of Chronic Disease - Hemolytic Anemia.

Using MATLAB program to build a knowledge base containing (175) image of the different types of anemia have been mentioned previously divided into 7 categories each category contains (15) images.



Name	Value	Min	Max
B	[0,0,255]	0	255
Bc	0	0	0
Bi	133	133	133
Black	[0,0,0]	0	0
Blackc	0	0	0
Blacki	3	3	3
Blackv	860	860	860
Bv	<1x14 doubl...	186	1144
CCV	<1x16 doubl...	0	250...
Color	<8x3 double>	0	255
Dist	[359.0711,35...	2.0...	439...
G	[0,255,0]	0	255
GB	[0,255,255]	0	255
GBc	0	0	0
GBi	0	0	0
GBv	[ ]		
GLDB	<1x30 struct>		
Gc	0	0	0
Gi	0	0	0

Fig 3: Part of the features of one of the images in the database

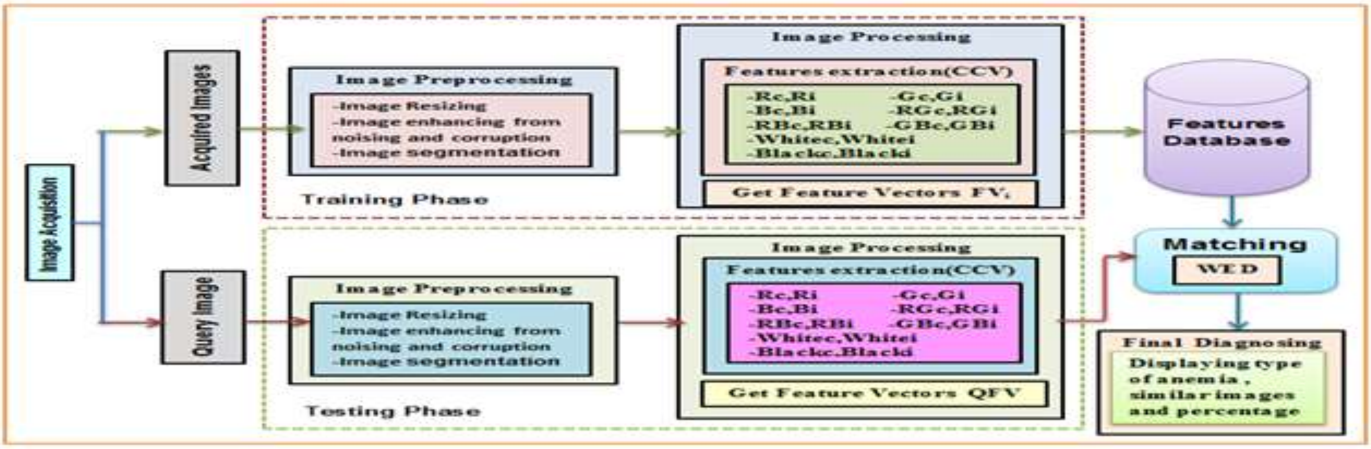


Fig 4: The framework of the proposed system

Second: Carrying out the processing process using the (ccv) method to extract the features:
This system aims to improve the images obtained using medical imaging devices or cameras [19].

In the training stage, color image features are extract using CCV. The images for the training process were acquired at the Dr. Ibrahim Abdel Hamid Ghazi laboratory for medical analysis in Minyat Al-Nasr and were also from medical references.

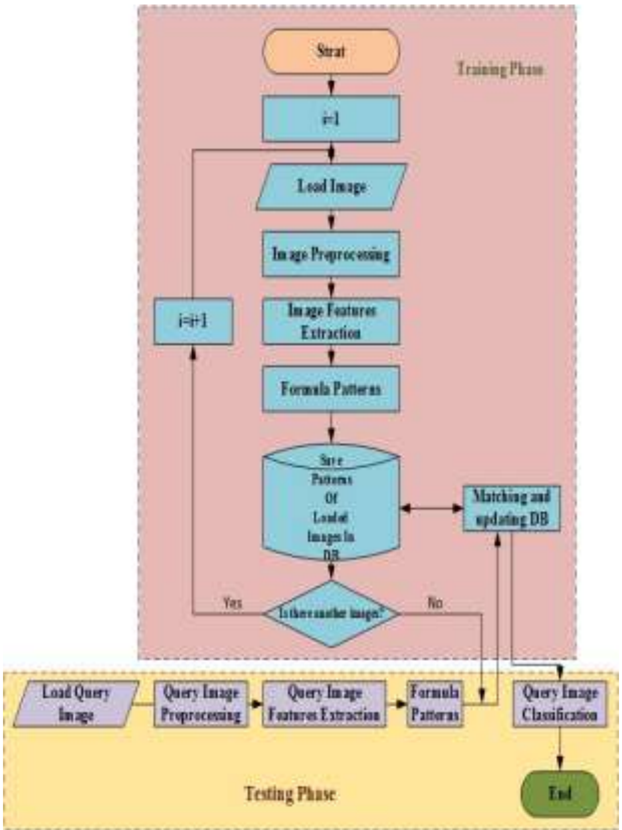
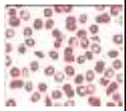
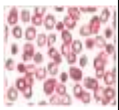


Fig5: Training Phase and Testing Phase.

The obtained images were shown to (5) experts in the field, to review its classification and divide it into (7) categories. Table2 shows examples of some types of anemia images.

In the image preprocessing stage, Images are resized and improved, which helps mitigate the effects of corruption and noise.

Table 2. Example type of anemia images.

No.	Images	Class name	No.	Images	Class name
1		Aplastic anemia	5		Iron deficiency anemia
2		Sickle cell anemia	6		Vitamin deficiency anemia
3		anemia hemolysis	7		Normal
4		bone marrow disease			

Figures (6,7) display the GUI for diagnosing some types of anemia

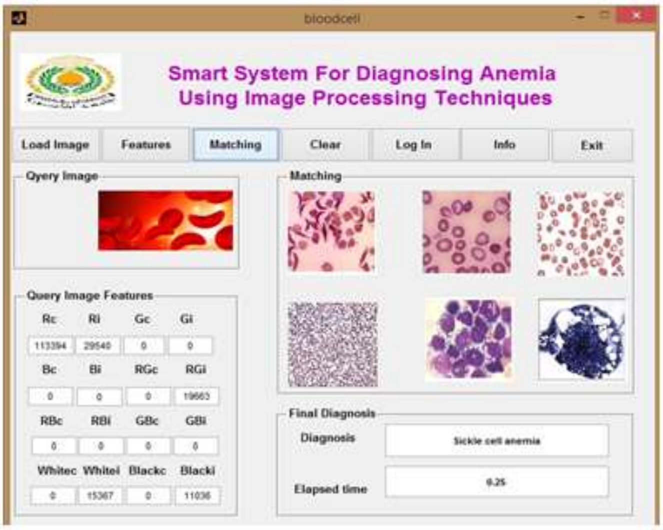


Fig6: GUI for diagnosing type of anemia images

Fig 7: The similar images

- Color Coherence Vector

To extract features from the images used Color Coherence Vector (CCV) represents this classification for each color in the image [20]. Number of adjacent and non-adjacent C_i pixels can be calculated as shown in equation (1).

The color coherence vector for the image Contains of: [21].

$$\langle(\alpha_1 + \beta_1)\rangle, \dots, \langle(\alpha_n + \beta_n)\rangle \quad (1)$$

Where:
a_n: number of coherent pixels
β_n: number of incoherent pixels
The similarity between these vectors will be obtained.[6]

$$D(a,b) = \sum_{i=0}^{n}(|a_{c_i} - b_{c_i}|) + (|a_{N_i} - b_{N_i}|) \quad (2)$$

Where:
C_i : number of coherent pixels colored with i.
N_i : number of incoherent pixels colored with i.

Features extracted are (16), the value of them is as follows: [Rc, Ri, Gc, Gi, Bc, Bi, RGc, RGi, RBc, RBi,GBc,GBi,Whitec, Whitei, Blackc, Blacki].

Field	Value	Min	Max
Rc	0	0	0
Ri	681	681	681
Gc	0	0	0
Gi	0	0	0
Bc	0	0	0
Bi	352	352	352
RGc	0	0	0
RGi	0	0	0
RBc	3794	3794	3794
RBi	6264	6264	6264
GBc	0	0	0
GBi	0	0	0

Fig8: Part of the query image features.

Third, a comparison process (matching):
- Matching Technique.

There are many methods used for pattern matching. The Weighted Euclidean Distance (WED) [22]. measure is the technique used in the proposed system.
- The WED measure is presented by This formula: [23]

$$d(v, v^k) = \sqrt{\sum_{i=1}^{n} p_i (v_i - v_i^k)^2} \quad (3)$$

$$\overline{v_i} = \frac{\sum_{k=1}^{N} v_i^k}{N} \quad (4)$$

$$P_i = \frac{N}{\sum_{K=1}^{N} (v_i^k - \overline{v}_i)^2} \quad (5)$$

Where:
v_i : to balance the differences in the dynamic scope.
P_i : weight added
K : is the Similar Image Index
N: the number of images in databases.

WED is given by producing the smallest value.[24]

The process of comparing the features extracted from the entered image and comparing them with the attributes are stored in the database.

- Final Decision

The steps to recognize anemia image are as follows:
1. Enter query image.
2. Find similar measurement between query image pattern and each image pattern in database using WED.
3. Sort similarity values in descending order.

The figure (9) shows the type of sickle cell anemia as a query image.

Fig9: Loading query image and extract features

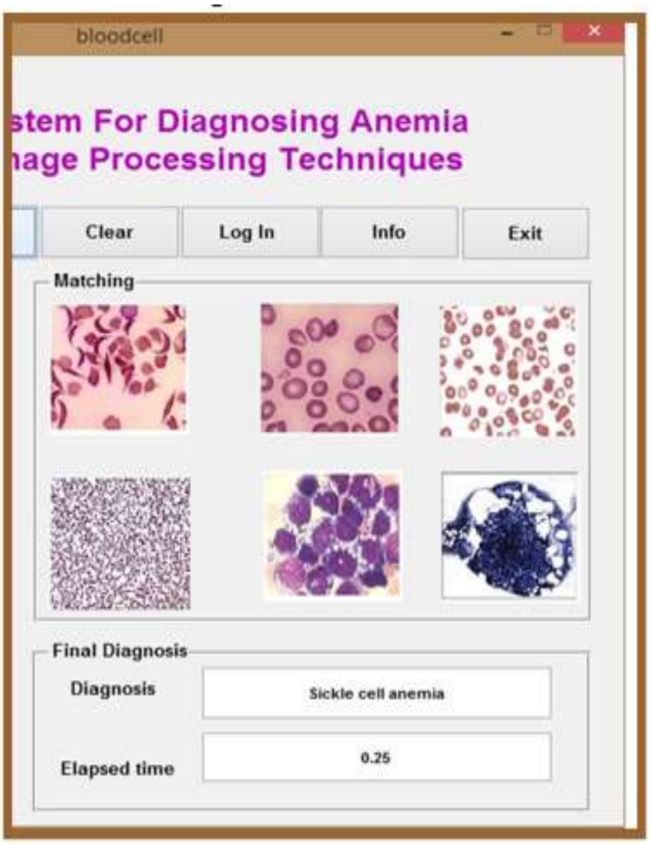


Fig 10: Retrieved Images by using Color Coherent Vector Feature Extraction Method

Figure 10 shows the resulted images due to using CCV feature extraction method of anemia image retrieval and Sort similar values in descending order.

- Fourth: Determining the appropriate type of anemia

The images were entered in specific categories. Each category had a specific number of images, if the image was found in a specific range of these categories, a result would be shown that this image belonged to this category and then the category name. But if the image was not within the scope of any category, the closest category would be suggested based on the similar ratio.

The figure (11) shows the result of the system, with the extraction of features, the similar images, the appropriate diagnosis, and the time elapsed for the operation.

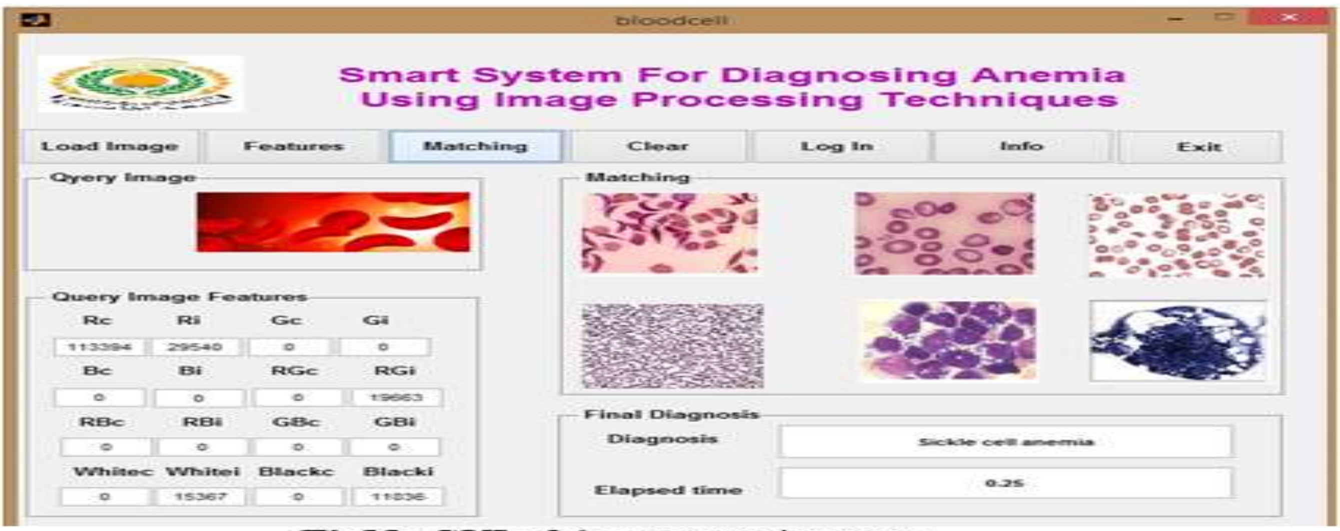


Fig11: GUI of the proposed system.

V. APPLICATIONS AND RESULTS

The proposed system was designed using MATLAB software. The CCV method is used to obtain a high similarity between the proposed classes. It can be used on a personal computer. To run the proposed system on a computer, MATLAB software is required. The graphic user interface of the proposed system is shown in Figure (13) and it is divided into four sections: (Inputting the image, extracting features, matching procedure and showing the diagnosis)

Database for testing uses (105) analysis samples of patients with the types of anemia mentioned earlier were provided to three medical analysis experts. Executing the diagnosis by the system and then comparing it with the expert diagnosis

Precision, Sensitivity, Specificity and F-measure were used. These measures are defined as: [25-26].

$$\text{Precision (P)} = \frac{\text{True positives}}{\text{True positives + False positives}} \quad 0 \le P \le 1 \quad (6)$$

$$\text{Sensitivity(S)} = \frac{\text{True positives}}{\text{True positives + False negatives}} \quad 0 \le R \le 1 \quad (7)$$

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives+ False positives}} \quad (8)$$

$$\text{F} - \text{measure} = 3 * \frac{\text{Precision} * \text{Sensitivity} * \text{Specificity}}{\text{Precision + Sensitivity + Specificity}} \quad (9)$$

Where:
True positives: correctly identified.
False positives: incorrectly identified.
False negatives: incorrectly identified.
True negatives: correctly identified

$$\text{Accuracy} = \frac{\text{sum of correct classifications}}{\text{Total number of classifications}} \text{x100} \quad (10)$$

$$\text{Accuracy} = \frac{100}{105} \text{x100= 95\%}$$

$$\text{Error Rate} = 1 \text{ - Accuracy} \quad (11)$$

Error Rate = 1 – 0.95 = 0:05%

Table 4. Display Confusion matrix for some types of anemia.

Type	Num	Aplastic	Sickle cell	hemolysis	bone marrow disease	Iron deficiency	Vitamin deficiency	Normal
Aplastic	15	4	0	1	0	0	0	0
Sickle cell	15	0	3	0	0	0	0	0
hemolysis	15	0	0	1	0	0	0	0
bone marrow	15	0	1	0	3	0	0	1
Iron deficiency	15	1	0	0	0	4	0	0
Vitamin deficiency	15	0	0	0	1	0	3	0
Normal	15	0	0	0	0	0	0	4

Table 3. Testing result

Disease Type	TP	TN	FP	FN	Precision	Sensitivity	Specificity	F-measure
Aplastic	**4**	**10**	**1**	**0**	**0.8**	**1**	**0.91**	**0.81**
Sickle cell	**3**	**12**	**0**	**0**	**1**	**1**	**1**	**1**
hemolysis	**1**	**14**	**0**	**0**	**1**	**1**	**1**	**1**
bone marrow disease	**3**	**12**	**0**	**2**	**1**	**0.6**	**1**	**0.69**
Iron deficiency	**4**	**10**	**1**	**0**	**0.8**	**1**	**0.91**	**0.81**
Vitamin deficiency	**3**	**12**	**0**	**1**	**1**	**0.75**	**1**	**0.82**
Normal	**4**	**11**	**0**	**0**	**1**	**1**	**1**	**1**
Aver					**0.94**	**0.91**	**0.97**	**0.88**

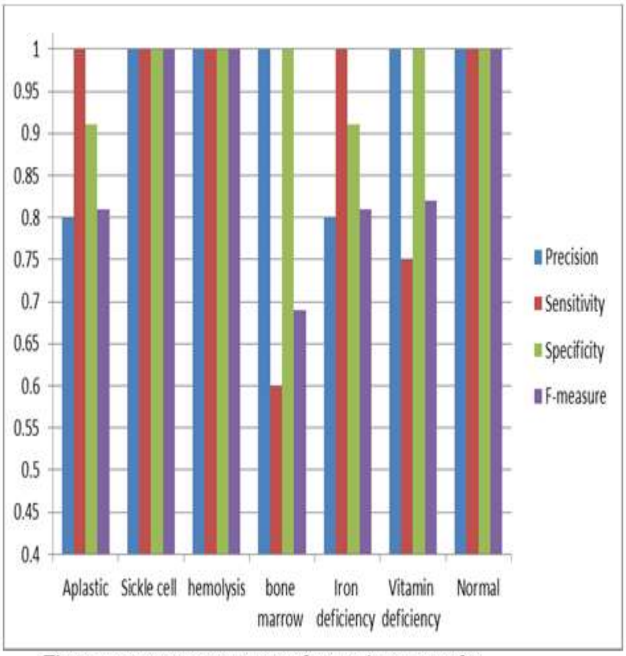


Fig 13: Precision, Sensitivity, Specificity and F-measure for classification diagnosing some anemia diseases.

This means that the proposed system can diagnose anemia diseases with satisfactory results.

VI. CONCLUSION

In this paper, it was presented how CCV is used to detect and diagnose some types of anemia. This can lead to a better classification of anemia and can be used for various purposes such as diagnosing blood diseases and assisting researchers in determining the type of anemia and its impact on a person's health. It can also help students in the Department of Home Economics to understand the topic of anemia classification and its relationship to therapeutic nutrition. Determining the size, shape and number of blood cells effectively affects the identification of blood diseases.

The proposed system can help confirm medical decisions. Based on the experimental results, The system accuracy rate is up to is 95%,

VII.RESULTS AND FUTURE WORK

This paper used the method of image retrieval in the diagnosis of anemia diseases, using the CCV method to build the proposed system. The results can be modified in future work by applying the proposed system to other types of anemia or different diseases. In future work the proposed system will be applied on different platforms such as mobiles and tablets.

VIII. REFERENCES

[1] **Manmohan Singhal, et al.,** "A research analysis on blood component usage and wastage in blood bank and blood component center", Academic Journals, vol.4, No.2,2013, pp23-28.

[2] **Mehmet Zahid Kocak, et al.,** Prevelance of anemia types and etiology in patients with anemia", Medicine Science, Vol.7,2018, pp 919-922.

[3] **Soundarya N, et al.**, " A review on anemia – types, causes, symptoms and their treatments", Journal of science and technology investigation (2456-8082), vol.1,2016, pp10-17.

[4] **Niloufar Mohammad khani,** " An introduction to the types of anemia", https://www.researchgate.net/publication/326572244**.2017**

[5] **Xuejian zhao, et al.,** " A Weighted Frequent Itemset Mining Algorithm for Intelligent Decision in Smart Systems ", IEEE Access, Vol. 6,2018, pp 29271 – 29282.

[6] **tarek mamdouh**, " Image Retrieval: Color Coherence Vector", **https://owlcation.com/ stem/ Image-Retrieval-Color-Coherence-Vector**, 2021.

[7] **S Y Veronica, M N Widyawati and Suryono**," Web-based rule-based system for early detection of anemia among pregnant mothers", IOP Conference of Materials Science and Engineering, ,2021,pp12-21, Indonesia.

[8] **M. Vadivel ,et al.**,"Classification of Sickle Cell Anemia Using Energy-Based KNN Classifier", Next Generation of Internet of Things,vol. 201,2021, pp289-293.

[9] **Bhavinkumar A. Patel and Ajay Parikh,** " Anemia Multi-label Classification Based on Problem Transformation Methods", Proceedings of Second International Conference on Computing , Communications, and Cyber-Security , vol. 203 ,2021, pp 627-639.

[10] **Shilpa A. Sanap , Meghana Nagori and Vivek Kshirsagar** , " Classification of Anemia Using Data Mining Techniques", International Conference on Swarm , Evolutionary , and Memetic Computing , vol.7077,2011, pp113-121.

[11] **Laith Alzubaidi, et al.,**"Classification of Red Blood Cells in Sickle Cell Anemia Using Deep Convolutional Neural Network", International Conference on Intelligent Systems Design and Applications, vol.940, 2018,pp 550-559.

[12] **Serhat kilicarslan, mete celik and şafak sahin**, " Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification", Biomedical Signal Processing and Control vol.63, 2021, p102231, Turkey.

[13] **Ning Zhang, et al.,** " Image retrieval of wool fabric Part II: based on low-level color features", Textile Research Journal, vol. 90, No. (7–8),2020, pp797–808.

[14] **Mahsa Moslehi, Felipe P. J. de Barros**, " Using Color Coherence Vectors to Evaluate the Performance of Hydrologic Data Assimilation" , water resources research ,vol.55 , Issue2 , 2019,pp 1717-1729.

[15] **Thoopsamut, et al.,** "Handwritten Signature Authentication using Color Coherence Vector and Signing Behavior", (2019), engineering village, vol.148384, 2019, pp 38-42.

[16] **Yibo Li**, Liu," Aerial Image Classification Using Color Coherence Vectors and Rotation Uniform Invariant LBP Descriptors", IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018, pp 653-656.

[17] **S. B. Mallikarjuna, et al.**, ," cnn based method for multi-type diseased arecanut image classification ", Malaysian Journal of Computer Science, Vol.34, No.3, 2021, pp.255-265.

[18] **Yousry Aql**, " **Challenges",** date of access:22/2/2013, **https://www.unicef.org/egypt/ar/nutrition**.

[19] **P. Ravi and A. Ashokkumar**, "Analysis of various image processing techniques," International Journal of Advanced Networking & Applications, vol. 8, No. 5,2017, pp. 86- 89.

[20] **Reshma Chaudhari and A. M. Patil**, " Content Based Image Retrieval Using Color and Shape Features", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol.1, No. 5, 2012, pp.387-388.

[21] **Greg Pass, Ramin Zabih and Justin Miller,** "Comparing Images Using Color Coherence Vectors", Proceedings of the fourth ACM international conference on Multimedia, vol.96,1996, pp. 66-67.

[22] **R. Mukundan and K. R. Ramakrishnan**, " Moment Functions in Image Analysis Theory and Applications", World Scientific Publishing, ISBN 981-02-3524-0,1998, pp. 81-85.

[23] **A. E. E. Elalfi, H. Fouda, and A. Atta,** "Developing an Intelligent decision support system for the diagnosis of some children's diseases", Computer Science Heart and Chest Surgery Department Mansoura University, vol. 151, No. 2,2016, pp. 32-38.

[24] **Ahmed Abd El-badie Abd Allah Kamel and Faten Abd El-Sattar Zahran El-Mougi,** " A Fuzzy Decision Support System for Diagnosis of Some Liver Diseases in Educational Medical Institutions", International Journal of Fuzzy Logic and Intelligent Systems, Vol. 20, No. 4, 2020, pp. 358-368.

[25] **Boris Neubert, et al.,** "Precision and Recall as Appearance Space Quality Measure for Simplified Aggregate Details", Eurographics Symposium on Rendering, Vol.210,2010, pp.55-70.

[26] **David M.W. Powers**, (2015) , " What the F-measure doesn't measure", https://www.researchgate.net/publication/273761233_What_the_F- measure doesn%27t_measure_ , available at date .10/5/2018 ,time.12.50 PM.