Automated Detection of White Blood Cells Cancer Diseases

Hend Mohamed, Rowan Omar, Nermeen Saeed,
Ali Essam, Nada Ayman, Taraggy Mohiy and Ashraf AbdelRaouf
Faculty of computer science
Misr International university
Email: hend1410317, rowan1405343, nermen1409132, ali1405090, nada.ayman, taraggy.ghanim, ashraf.raouf{@miuegypt.edu.eg}

***Abstract*—Automated diagnosis of white blood cells cancer diseases such as Leukemia and Myeloma is a challenging biomedical research topic. Our approach presents for the first time a new state of the art application that assists in diagnosing the white blood cells diseases. we divide these diseases into two categories, each category includes similar symptoms diseases that may confuse in diagnosing. Based on the doctor's selection, one of two approaches is implemented. Each approach is applied on one of the two diseases category by computing different features. Finally, Random Forest classifier is applied for final decision. The proposed approach aims to early discovery of white blood cells cancer, reduce the misdiagnosis cases in addition to improve the system learning methodology. Moreover, allowing the experts only to have the final tuning on the result obtained from the system. The proposed approach achieved an accuracy of 93% in the first category and 95% in the second category.**

***Keywords*—white blood cells disorder, classification, Deep Learning, Random Forest**

I. Introduction

According to the World Health Organization (WHO) [1], Cancer is consider as the second leading cause of death in the world and has claimed the lives of 8.8 million people in 2015, attributed to the death of nearly one in six deaths worldwide. According to WHO [1] Low-income and middle-income countries accounted for almost 70% of cancer deaths. Moreover, Egypt is ranked as the 19^{th} worldwide among 176 countries in leukemia death rates. Discovering those diseases in the early stages highly affects the treatment period. In addition, some of the diseases' sub-types are really confusing to the doctors. Nowadays, there is a great tendency for diagnostic pathology to heavily rely on automated systems which can aid in the diagnosis [2].

Image processing is concerning with digital images to extract useful information. It is involved in different other topics such as layout land use [3], Character recognition [4][5], coin recognition [6], Medical imaging [7][8]. Medical imaging uses the information extracted from digital image to enhance the diagnostic of different diseases.

White blood cells cancer diseases; Leukemia and Myeloma, threaten people's life nowadays. Leukemia is found when the bone marrow produces abnormal white blood cells, which don't function properly [9]. It may be either acute or chronic. Acute Myeloid Leukemia (AML) is sub categorized to (M0, M1, M2, M3, M4, M5, M6, M7). Acute lymphoblastic leukemia (ALL) is sub categorized to (L1, L2, L3). Myeloma [10] is another type of cancer that develops from cells in the bone marrow called plasma cells.

This paper introduced for the first time a hybrid automated system to facilitate the diagnosis of different white blood cells cancer diseases; Leukemia with its sub-types (AML, ALL) and Myeloma. Two main approaches are conducted for the sake of eliminating the confusion while discriminating between some of the sub-types. A new combination of ratio features are conducted for each approach. We considered taking ratio features because it is invariant to scaling. Finally, we tried to make the system as an expert system through improving the learning efficiency by allowing the system to learn from the misdiagnosed input tests in which the experts reclassify the disease with the new label.

The paper is organized as follow, section II discuss similar related research. The proposed approach through the two different category is presented in Section III. Section IV analyses and discusses the experimental results. Finally sections V presents our conclusion and future work.

II. Related Work

Some research were proposed to differentiate between:

1) AML and ALL diseases: Karthikeyan and Poornima [11] proposed an approach for detection of Leukemia in blood at early stages. They have used adaptive median filter for noise removal and adaptive Histogram Equalization for contrast enhancement in preprocessing stage. They applied k-means and Fuzzy c-means clustering for segmentation. They computed statistical, textural and geometrical features and applied Support Vector Machine (SVM) for classification. Their approach achieved 90% with Fuzzy c-means and 83% with k-means using Fuzzy Logic: Intelligence, Control, and Information dataset [12]. Another research by Mohapatra et al. [13] proposed a Fuzzy based Blood Image Segmentation for Automated Leukemia Detection. They applied selective median filtering followed by unsharp masking in preprocessing. In segmentation, they used improved version of fuzzy clustering technique viz. Gustafson Kessel clustering [14] followed by nearest neighbor classification in L*a*b* color space (L* for lightness, a* for rednessgreenness axis, and b* a yellownessblueness axis) [15]. The computed features are two novel shape features; Hausdorff Dimension and contour signature. Support Vector Machine (SVM) is employed for classification and they achieved 93% on a database of 108

blood smear images of size 512 x 512 pixels.

2) AML detection only or ALL detection only: An approach by Agaian et al. [16] proposed a simple technique that automatically detects and segments AML in blood smears. Segmentation was done in the CIELAB Color space by K-Means clustering algorithm. Hausdorff Dimension features were computed using the box counting method and Local Binary Pattern (LBP). Classification achieved 98% accuracy using the Support Vector Machine (SVM) on American Society of Hematology (ASH) for Leukemia dataset [17]. This dataset comprised of 80 images40 from AML patients and 40 from non-AML patients. The image size used for their classification was 184 x 138 pixels. Another system proposed by Bhattacharjee and Saini [18] for Acute Lymphoblastic Leukemia detection using Watershed Transformation Technique. They applied contrast enhancement and quality adjustment for enhancing images before segmentation. In segmentation they used watershed algorithm, isolating the blood cell and the cell nucleus. They computed area, perimeter, circularity and form factor features. Gaussian Mixture Models (GMM) and Binary Search Tree (BST) were applied for classification. GMM achieved 93% while BSTe achieved 86%. They applied their approach on 150 lymphocytic cells images (30 normal cells & 120 blast cells) accessed from ALL-IDB1 & ALLIDB2 datasets [17].

3) Detection of (AML) sub-types: Another proposed approach by Sarrafzadeh et al. [19] focused mainly on differentiate between M2, M3 and M5 sub-types to evaluate their introduced method. The approach was applied in the L*a*b* color space. Segmentation is performed using K-means clustering to segregate leukocytes from other blood components. Texture and shape features are extracted in order to be classified using Discriminative Dictionary Learning (DDL). They achieved 97.53% accuracy on Medical Image & Signal Processing Research Center (MISP) dataset [20]. They used a dataset composed of 27 microscopic images of three sub-types of AML; 9 AML-M2, 10 AML-M3 and 8 AML-M5.

III. Approach

Our proposed system composed of two contributory approaches. One differentiate between M5 Acute Myeloid Leukemia (AML), L1 and L2 Acute lymphoblastic leukemia(ALL), while the second differentiate between the remaining sub categories. We grouped our concerned diseases into two separate sets due to similar visual features that actually confuse doctors and may cause misclassification. According to the doctor expectation of the input blood sample, the system goes through one of the two proposed approaches. A different set of features is computed per approach. We used supervised classification, as we trained our system with labeled data of the diseases' sub-types samples. The input blood test sample is then classified under the supervision of the matching similarities between it and the trained data. Cascading recognition process by two approaches increase the overall system efficiency of discriminating between the diseases as stated in the experiment section. The system overview is shown in Fig.1.

A. Preprocessing & Segmentation

This phase is applied on our training and testing images. The purpose of white blood cell segmentation is to clearly extract relevant object; whole cell, from its relative background.

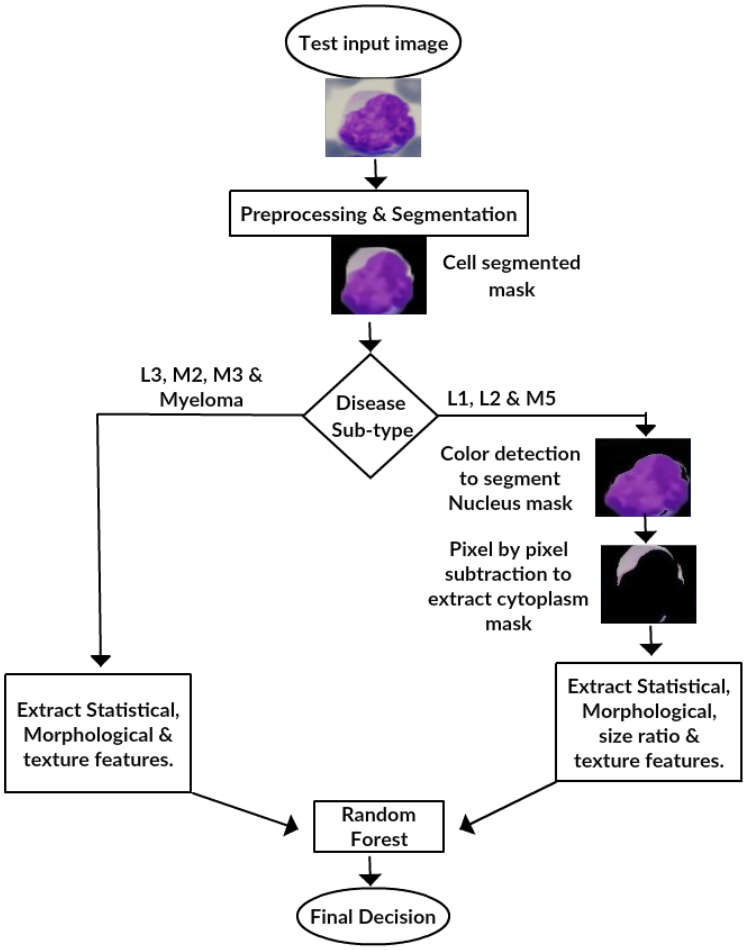

Fig. 1. The proposed approach flowchart

Furthermore, we separate the cell into nucleus and cytoplasm .

1) Preprocessing: In this stage we prepare the blood sample image for the segmentation process by converting our input images from RGB color space to YCbCr space [21]. Our choice to the YCBCR color space (Y: Luminance, CB: Blue Value, CR: Red Value) was due to the reddish and bluish colors of our blood samples. After converting images to YCbCr space, the extracted Cb and Cr coefficients are used for cell segmentation process. Sample input image before and after conversion is shown in fig. 2.

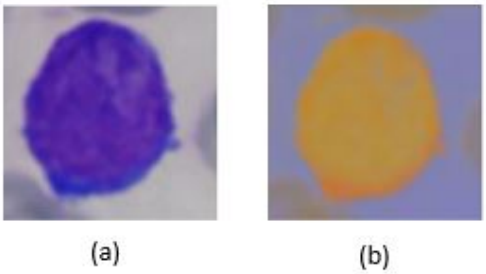

Fig. 2. Cell input images of AML-M2 before and after Preprocessing (a)RGB Color Space, (b)YCBCR Color Space

2) Cell Segmentation: The purpose of this stage is to segment the whole cell from the relative background as shown in Fig.3. The extracted Cb and Cr coefficients from our training images during preprocessing stage are now used to build a Gaussian Distribution [22] as shown in equations 1, 2, 3, 4.

$$bmean = mean(cb) \tag{1}$$

Where cb is the row vector containing all Cb coefficients obtained from our training images, and bmean is the blue mean of this vector.

$$rmean = mean(cr) \tag{2}$$

Where cr is the row vector containing all Cb coefficients obtained from our training images, and rmean is the red mean of this vector.

$$brcov = cov(cb, cr) \tag{3}$$

Where brcov is the co-variance of the two row vectors cb and cr.The result is a 2x2 matrix.

$$magCov = (brcov(1,1) * brcov(2,2) - brcov(2,1) * brcov(1,2)) \tag{4}$$

Where magCov is the magnitude of the brcov.

In the testing phase the Guassian distribution is applied on the input test image to accomplish the segmentation stage as shown in equations 5, 6.

$$x = [(cb - bmean), (cr - rmean)] \tag{5}$$

$$f(x) = \frac{e^{-0.5 * x * brcov^{-1} * x'}}{2 * \pi * magCov} \tag{6}$$

This Gaussian distribution is applied to test images in the YCbCr space to extract our valuable pixels that are most probable included to our regions of interest ROI. After applying our defined distribution, normalization is applied followed by adaptive threshold algorithm [23].

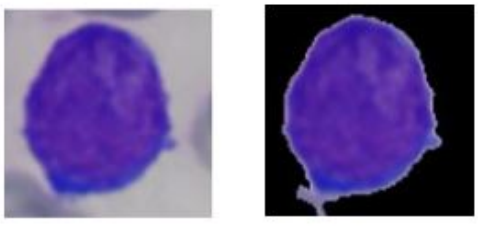

Fig. 3. (a)Cell original RGB image of AML-M2 before segmentation, (b)Cell (a) after Segmentation

3) Nucleus & Cytoplasm Segmentation: The result from cell segmentation is a mask containing only the cell. Color detection technique [24] is applied on the cell mask with specified range of colors to segment nucleus mask [25]. By simple pixel to pixel subtraction of these two masks we can easily extract an accurate mask for the cytoplasm as shown in Fig. 4

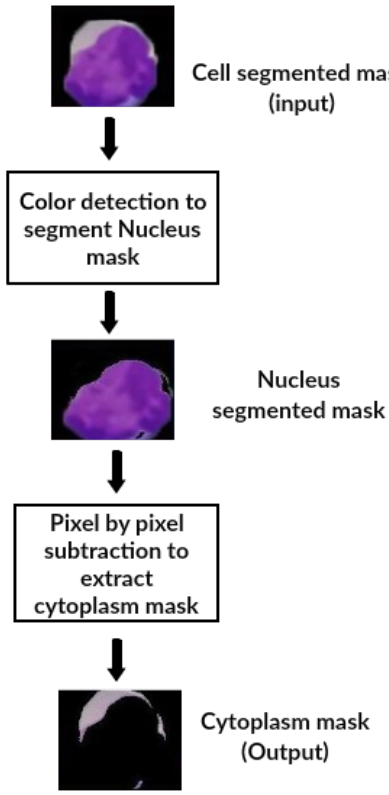

Fig. 4. Nucleus and Cytoplasm Segmentation

B. Feature Extraction

This phase is applied on segmented images resulted from preprocessing and segmentation process.The differentiation between multiple types of both ALL, AML and Myeloma require computing different types of features to compensate visual similarities. In our proposed approach, we computed morphological, statistical, size ratio [26] and texture features [27]. According to the running approach, different set of features are calculated. The first approach concerns L1, L2, M5 subcategories while the other concerns L3, M2, M3 and Myeloma. Based on the doctors' decision one of the two approaches will be followed. We considered ratio features because they are invariant to scaling as discussed briefly in the following subsections.

1) Morphological Features: These features represent shape of the cell and its dimension [26]. The calculated features are area to perimeter ratio, circularity, elongation, major to minor axis length ratio, extent and solidity.

a) Area to perimeter ratio: It is the ratio between the actual number of pixels in the Region Of Interest (ROI) and the distance between each adjoining pair of pixels around the

border of the ROI.

$$AreaToPerimeterRatio = \frac{Area}{Perimeter} \quad (7)$$

b) Circularity: This feature measures the complexity of the perimeter of the circular object.

$$Circularity = \frac{Perimeter}{(4 * Area * pi)} \quad (8)$$

c) Elongation: It is the ratio between length of the smallest rectangle containing the ROI and width of the smallest rectangle containing the ROI. It is also known as the growth in one direction of the ROI.

$$Elongation = \frac{LSR}{WSR} \quad (9)$$

where LSR is the length of the smallest rectangle containing the ROI and WSR is the width of the smallest rectangle containing the ROI.

d) Major to minor axis length ratio: It is the ratio between the major axis of the ellipse containing the ROI and the minor axis of the ellipse containing the ROI.

$$MajorToMinorAxisLengthRatio = \frac{MajorAxisLength}{MinorAxisLength} \quad (10)$$

e) Extent: It is the proportion of ROI area to the area of its bounding rectangle.

$$Extent = \frac{Area}{(Width * Length)} \quad (11)$$

f) Solidity: It is the proportion of ROI area to area of its convex hull.

$$Solidity = \frac{Area}{Convexarea} \quad (12)$$

2) Statistical Features: These features also concern cell shape information but from different perspective [26]. The calculated features are the following

a) Mode: It is defined as most frequent value of the pixels intensity of the ROI.

b) Mean: It is the average value of the pixels intensity of the ROI.

c) Standard deviation: Standard deviation is a value represent how much pixels intensity differ from the mean of pixels intensities of the ROI.

d) Variance: Variance value of the pixels intensity of the ROI.

e) Sum: Sum of the pixels intensities of the ROI.

f) Gradient: Angles' gradient is calculated by Canny edge detection.

3) Size ratio features: Additional set of features are extracted based on the Segmentation of nucleus and cytoplasm as introduced in [26].

a) Nucleus cytoplasm area: It is the ratio of the area between the nucleus and the cytoplasm.

$$NucleusToCytoplasmArea = \frac{NucleusArea}{CytoplasmArea} \quad (13)$$

b) Nucleus cell area: It is the ratio of the area between the nucleus and the cell.

$$NucleusToCellArea = \frac{NucleusArea}{CellArea} \quad (14)$$

c) Nucleus cell perimeter: It is the ratio of the perimeter between the nucleus and the cell.

$$NucleusToCellPerimeter = \frac{NucleusPerimeter}{CellPerimeter} \quad (15)$$

4) Texture Features: These features concern details in the cell like holes and granules. We implemented Haralicks features [27]. It is a set of 14 texture features calculated from the gray level co-occurrence matrix using 4 directions of adjacency. These features are angular second moment, contrast, correlation, variance, inverse different moment, sum average, sum variance, sum entropy, entropy, difference entropy, difference variance, measure of correlation 1, measure of correlation 2 and maximum correlation coefficient.

C. Classification

Random forest algorithm [28], [29] is a supervised classification algorithm that constructs a forest with several decision trees. Highest accuracy results are achieved with the higher number of trees. Random forest algorithm achieved successes in medical field [30] as its one of the most powerful algorithms that is widely used in different applications. It has many advantages as it can be used in different classification problems such as banking, stock market and E-commerce, it can be used for both classification and regression and it performs feature selection to only extracts the crucial features.

In our proposed method, Random Forest Classifier is used for the two main categories of the system. Random Forest

TABLE I. COMPARISON BETWEEN OUR EXPERIMENTS BEFORE THE SPLITTING APPROACH

Exp. No.	Features	Classifier	L1	L2	L3	M3	M5	Myeloma	Train Accuracy	Test Accuracy
1	Statistical	Random Forest	untested	untested	100%	100%	100%	80%	97.5	90%
2	Morphological	Random Forest	untested	untested	60%	60%	0%	40%	97.5	68%
3	Statistical & Morphological	Random Forest	untested	untested	100%	100%	80%	80%	100%	90%
4	Size ratio, (N-Cell area, N-Ctyto area, & N-Cell perimeter)	Random Forest	untested	untested	100%	0%	80%	60%	100%	60%
5	Size ratio, (N-Cell area, N-Ctyto area, & N-Cell perimeter)	Random Forest	untested	untested	100%	untested	80%	60%	100%	80%
6	Statistical, Morphological & size ratio	Random Forest	untested	untested	100%	100%	80%	80%	100%	90%
7	Statistical, Morphological & size ratio	Logistic regression	untested	untested	80%	100%	80%	40%	92.5%	75%
8	Statistical,Morphological, Texture & size ratio	Random Forest	untested	untested	100%	100%	100%	60%	100%	90%
9	Statistical, Morphological &Texture	Random Forest	100%	80%	40%	80%	100%	60%	100%	76%
10	Morphological & Texture	Random Forest	80%	80%	60%	100%	100%	60%	98%	80%
11	Statistical, Morphological & size ratio	Naive Bayes	untested	untested	100%	100%	100%	60%	87.5%	90%
12	Morphological & Texture	Naive Bayes	40%	60%	80%	80%	80%	60%	78%	63%
13	Morphological & Texture	Logistic Regression	60%	40%	40%	100%	80%	60%	100%	63%

TABLE II. COMPARISON BETWEEN OUR EXPERIMENTS AFTER THE SPLITTING APPROACH

Exp. No.	Features	Classifier	L1	L2	L3	M2	M3	M5	Myeloma	Train Accuracy	Test Accuracy
1	Statistical, Morphological, Texture & size ratio	Random Forest	80%	80%	80%	untested	100%	80%	60%	100%	80%
2	Statistical, Morphological, Texture & size ratio	Random Forest	100%	100%	untested	untested	untested	untested	untested	100%	100%
3	Statistical, Morphological, Texture & size ratio	Random Forest	100%	100%	untested	untested	untested	80%	untested	100%	**93%**
4	Statistical, Morphological & Texture	Random Forest	100%	60%	untested	untested	untested	100%	untested	100%	86%
5	Statistical, Morphological & Texture	Random Forest	untested	untested	100%	100%	100%	untested	80%	100%	**95%**
6	Statistical, Morphological & Texture	Random Forest	untested	untested	100%	100%	100%	100%	60%	100%	92%

classifier is the best classifier that is able to differentiate between different types and the one which gives us the highest accuracy as stated in our experiments table I and II. Also, the architecture that this classifier is based on fits our problem as we have three parent disease classes including ALL, AML and Myeloma and each one has many sub-classes as their sub-types.

D. Expertising and relearning

Medical Automated recognition systems are very sensitive systems. It acts as a preliminary decision maker that influence doctors' final decisions. During The Classification process, a misdiagnosis may occur. We introduced the idea of system expertising to be trustable enough to the doctors in the diagnosis process. Experts are allowed to have the final decision either to accept the result from the proposed system or reclassify it with the correct label based on their medical experience. The system is capable of relearning from the misdiagnosed cases through feeding the system again with those newly classified samples by the expert doctors by repeating the training process.

IV. EXPERIMENTAL RESULTS

The proposed algorithm was tested and trained on a dataset that contains 105 cell images of three types of diseases including ALL, AML and Myeloma; 15 AML-M2, 15 AML-M3, 15 AML-M5, 15 ALL-L1, 15 ALL-L2, 15 ALL-L3 and 15 Myeloma. Images are in PNG format with a size of 100x100. Datasets were obtained from Medical Image and Signal Processing Research Center (MISP) available at [20] and Myeloma dataset is obtained from [31] Test image can be classified into whether it's M2, M3, M5, L1, L2, L3 or Myeloma. Fig.5 shows sample cell images of our dataset.

In our experiments, different combinations of extracted fea-

TABLE III. COMPARISON WITH SIMILAR SYSTEMS

	Detection	Segmentation	Classifier	Accuracy	Dataset	Learning System
Our System	ALL(L1,L2 and L3) AML(M2,M3 and M5) Myeloma	Gaussian Distribution	Random Forest	93% (L1,L2,M5) 95% (L3,M2,M3,Myeloma)	105 images (70 training, 35 testing)	Present
[11]	ALL and AML.	K-means clustering Fuzzy c means clustering	SVM	90% Fuzzy c means 83% K-means	In Fuzzy Logic: Intelligence Control and Information Book	Doesn't present
[13]	ALL	Fuzzy clustering	SVM	93%	108 blood smear images	Doesn't present
[16]	AML	K-means clustering.	SVM	93.5% without LBP 97.5% with LBP	80 images (40: AML,40 normal)	Doesn't present
[18]	ALL	Watershed Algorithm.	GMM, Binary search tree	93% with GMM 86% with Binary Search	150 cell images (30 normal,120 blast) 75% training 25% testing	Doesn't present

tures and classifiers were tested on different types of diseases. We applied many alternatives concerning core algorithm to choose the best that fits our problem. Those alternatives include different classifiers such as support vector machine [25], [32], Random Forest [29], [28], [30], Naive bayes [33] and Logistic regression [34]. Random forest achieved very promising results as we have three parent disease classes and each one has many sub-classes as their sub-types. It was very challenging to differentiate between L1, L2 and M5 as they have very similar visual features that the classifier couldn't recognize them easily. As mentioned before in our approach, we divided these sub-types into two sets and decreased confusion as some of the diseases seemed to be visually similar.

In our Experiments we implemented two approaches in which the system can go through based on the doctor's decision. Each approach concerns set of diseases. We split the diseases into two different sets as shown in table I. The first approach differentiates between L1, L2, M5 which achieved 93% accuracy with statistical, morphological, size ratio and texture features, while the other approach differentiates between L3, M2, M3 and Myeloma which achieved 95% accuracy with statistical, morphological and texture features. In tables I and II, there are some sub-types not included in our experiments and will be targeted in our future work. We applied different sets of features until we reached to the best combination of sub-types and features that gives the best accuracy. The accuracy achieved by our proposed splitting approach is shown in table II. The system overall accuracy achieved is 94.3%. The splitting approach enhanced the accuracy from 80% as shown in table I to 94.3% as shown in table II.

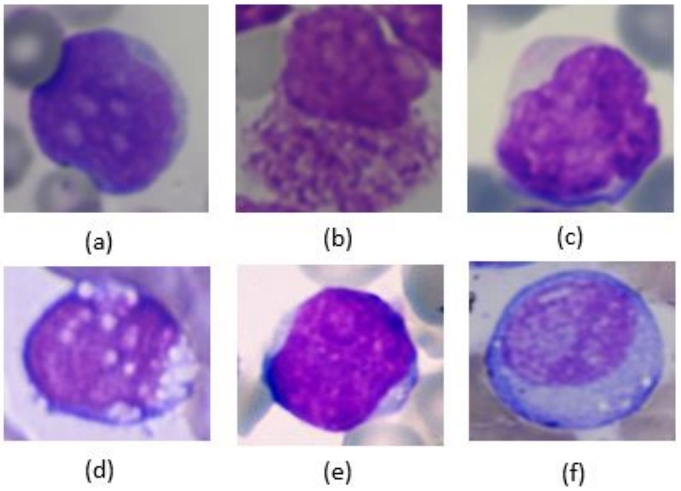

Fig. 5. Sample images of ALL, AML & Myeloma, (a)M2, (b)M3, (c)M5, (d)L3, (e)L2, (f)Myeloma

A comparison between our system and other similar systems is summarized in table III. The comparison includes the differences between the applied techniques and the achieved results. Similar systems in our case were divided into two categories, the first category was concerned with detecting one type at a time without taking into consideration the sub-types of this disease. These systems determine whether input sample is blast or not. The second category was concerned with detecting one type with its sub-types. Our main contribution is taking into consideration many types of diseases together with their sub-types. We also applied new segmenting technique based on Gaussian distribution to extract the cell from the whole image while other systems used different techniques like k-means and watershed. Our implementation enables experts to re-train the system and re-learn in order to increase the accuracy of the proposed system.

V. CONCLUSION

In this paper we propose the design, development and evaluation of an automated system to accurately detect white blood cells cancer diseases. It detects types and sub-types of Leukemia (ALL and AML) and Myeloma. It is a recognition system applied on acquired blood microscopic images then performs preprocessing, segmentation, feature extraction and classification. The proposed solution converts images to YCBCR color space and construct Gaussian distribution of

CB and CR values. Statistical, texture, size ratio and morphological features are then computed to train classifier. Unlike existing systems, our system has the ability of learning from misclassified tests to enhance the future accuracy of the system. Random Forest classifier is the best classifier that is able to differentiate between different types and the one which gives us the best accuracy. The system achieved 94.3 % accuracy in detecting and classifying types and sub-types.

As our next step, we aim to detect more types of white blood cells cancer diseases to build an overall system for white blood cells diseases.

REFERENCES

[1] K. M. Sensing, Cancer.
URL http://www.who.int/mediacentre/factsheets/fs297/en.html

[2] J. B. Henry, J. P. AuBuchon, Clinical diagnosis and management by laboratory methods, Archives of Pathology and Laboratory Medicine 121 (9) (1997) 1016.

[3] E. Sarhan, E. Khalifa, A. M. Nabil, Post classification using cellular automata for landsat images in developing countries, in: International Conference on Image Information Processing (ICIIP), Shimla, India, IEEE, 2011, pp. 1–4.

[4] A. AbdelRaouf, C. A. Higgins, T. Pridmore, M. I. Khalil, Arabic character recognition using a haar cascade classifier approach (hcc), Pattern Analysis and Applications 19 (2) (2016) 411–426.

[5] A. AbdelRaouf, C. A. Higgins, T. P. Pridmore, M. I. Khalil, Arabic corpus enhancement using a new lexicon/stemming algorithm., in: 2nd International Conference on Pattern Recognition Applications and methods (ICPRAM), Barcelona, Spain, 2013, pp. 435–440.

[6] T. M. Ghanem, M. N. Moustafa, H. I. Shahein, Gabor wavelet based automatic coin classsification, in: 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, IEEE, 2009, pp. 305–308.

[7] S. Sameh, M. A. Azim, A. AbdelRaouf, Narrowed coronary artery detection and classification using angiographic scans, in: 12th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, IEEE, 2017, pp. 73–79.

[8] S. Moataz, M. A. Azim, A. AbdelRaouf, Automated fish signals fusion detection for chronic myeloid leukemia diagnosis, in: 26th International Conference on Computer Theory and Applications (ICCTA), Alexandria, Egypt, IEEE, 2016.

[9] M. F. for Medical Education, Research, Leukemia.
URL https://www.mayoclinic.org/diseases-conditions/leukemia/symptoms-causes/syc-20374373.html

[10] What is multiple myeloma?
URL https://www.cancer.org/cancer/multiple-myeloma/about/what-is-multiple-myeloma.html

[11] T. Karthikeyan, N. Poornima, Microscopic image segmentation using fuzzy c means for leukemia diagnosis, Leukemia 4 (1).

[12] J. Yen, R. Langari, Fuzzy logic: intelligence, control, and information, Vol. 1, Prentice Hall Upper Saddle River, NJ, 1999.

[13] S. Mohapatra, S. S. Samanta, D. Patra, S. Satpathi, Fuzzy based blood image segmentation for automated leukemia detection, in: Devices and Communications (ICDeCom), 2011 International Conference on, IEEE, 2011, pp. 1–5.

[14] D. Graves, W. Pedrycz, Fuzzy c-means, gustafson-kessel fcm, and kernel-based fcm: A comparative study, in: Analysis and Design of Intelligent Systems using Soft Computing Techniques, Springer, 2007, pp. 140–149.

[15] K. M. Sensing, Identifying color differences using l*a*b* or l*c*h* coordinates.
URL https://sensing.konicaminolta.us/blog/identifying-color-differences-using-l-a-b-or-l-c-h-coordinates/

[16] S. Agaian, M. Madhukar, A. T. Chronopoulos, Automated screening system for acute myelogenous leukemia detection in blood microscopic images, IEEE Systems journal 8 (3) (2014) 995–1004.

[17] R. D. Labati, V. Piuri, F. Scotti, All-idb: The acute lymphoblastic leukemia image database for image processing, in: Image processing (ICIP), 2011 18th IEEE international conference on, IEEE, 2011, pp. 2045–2048.

[18] R. Bhattacharjee, L. M. Saini, Detection of acute lymphoblastic leukemia using watershed transformation technique, in: Signal Processing, Computing and Control (ISPCC), 2015 International Conference on, IEEE, 2015, pp. 383–386.

[19] O. Sarrafzadeh, H. Rabbani, A. M. Dehnavi, A. Talebi, Detecting different sub-types of acute myelogenous leukemia using dictionary learning and sparse representation, in: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 3339–3343.

[20] Medical image and signal processing research center.
URL http://misp.mui.ac.ir/en

[21] T. Acharya, P.-S. Tsai, JPEG2000 standard for image compression concepts, algorithms and VLSI architectures, Wiley-Interscience, 2005.

[22] K. Krishnamoorthy, Handbook of statistical distributions with applications, CRC Press, 2016.

[23] A. Elmoataz, Image and signal processing 5th International Conference, ICISP 2012, Agadir, Morocco, June 28-30, 2012. Proceedings, Springer, 2012.

[24] R. Lukac, K. Plataniotis, Color Image Processing: Methods and Applications, Image Processing Series, CRC Press, 2006.

[25] E. A. Mohammed, B. H. Far, C. Naugler, M. M. Mohamed, Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation, in: e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on, IEEE, 2013, pp. 622–626.

[26] C. Reta, L. Altamirano, J. A. Gonzalez, R. Diaz-Hernandez, H. Peregrina, I. Olmos, J. E. Alonso, R. Lobato, Segmentation and classification of bone marrow cells images using contextual information for medical diagnosis of acute leukemias, PloS one 10 (6) (2015) e0130805.

[27] R. M. Haralick, Statistical and structural approaches to texture, Proceedings of the IEEE 67 (5) (1979) 786–804.

[28] A. Liaw, M. Wiener, Classification and regression by random forest, R News 2 (3) (2002) 18–22.

[29] Y. Pavlov, Random Forests, VSP, 2000.

[30] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC bioinformatics 10 (1) (2009) 213.

[31] G. Sites, Bone marrow microscopic data - dr hossein rabbani.
URL https://sites.google.com/site/hosseinrabbanikhorasgani/datasets-1

[32] I. Steinwart, A. Christmann, Support Vector Machines, Information Science and Statistics, Springer New York, 2008.

[33] M. Kirk, Thoughtful Machine Learning: A Test-Driven Approach, O'Reilly Media, 2014.

[34] D. Kleinbaum, M. Klein, Logistic Regression: A Self-Learning Text, Statistics for Biology and Health, Springer New York, 2010.