ZANCO Journal of Pure and Applied Sciences
The official scientific journal of Salahaddin University-Erbil
https://zancojournals.su.edu.krd/index.php/JPAS

ISSN (print):2218-0230**, ISSN (online):** 2412-3986**, DOI: http://dx.doi.org/10.21271/zjpas**

RESEARCH PAPER

Automated segmentation of Acute Lymphocytic Leukemia (ALL) subtypes by the combination of color space conversion and K-means cluster

[1]Hersh Muhsin Osman *, [2]Sardar Pirkhider Yaba

[1,2] Department of Physics, College of Education, Salahaddin University-Erbil, 44001, Erbil, Kurdistan, Iraq

A B S T R A C T:

Leukemia is blood cancer, and it is one of the most common and deadly causes of death in the world. Morphologically, Leukemia cells are classified into three types of L1, L2, and L3 by the French- American- British (FAB) classification. A new method of automatically segmenting blast cells from microscopic blood smear images proposed in this research. This study proposes significant pre-processing to obtain high segmentation performance and presents a new combination of image processing approaches. specially, the five color spaces selected with K-means cluster to segment subtypes of Acute Lymphocytic Leukemia. The majority of the components of the performance color space were chosen based on their similarity with ground truth image through using five evaluation parameters. The proposed codes for Acute Lymphocytic Leukemia subtype accurate segmentation applied on local and public datasets (427 images). The best color space type was YIQ which had 87% performance for the public dataset with segmentation evaluation 96.24% for dice parameter

KEY WORDS: Acute lymphoblastic leukemia (ALL), color Space conversion, K-means clustering;
DOI: http://dx.doi.org/10.21271/ZJPAS.34.3.2
ZJPAS (2022) , 34(3);11-20 .

1. INTRODUCTION:

Acute leukemia is a blood-forming cell cancer that is characterized by uncontrollable and immature leukocyte production (WBCs). According to the World Cancer Statistics report, 19.3 million new cancer patients were diagnosed in 2020. with leukemia accounting for 2.5 percent of those diagnosed and 3.1 percent of neoplastic fatalities, with the mortality rate ranking 10th (Siegel et al., 2020). Annually in the United States, the report indicated that 3,000 to 4,000 people develop acute lymphoblastic leukemia (ALL), two-thirds of whom are children (Philip et al., 2021). Also, the rate of leukemia in the Erbil Governorate (Kurdistan/ Iraq) due to other types of cancers was 5.27%, according to Awat center of cancer statistics in Erbil during 2015-2020. Among them, 4.01% is due to acute leukemia (Karwan et al., 2021).

* **Corresponding Author:**
Hersh Muhsin Osman
E-mail: hershmuhsinphysic@gmail.com
Article History:
Received: 23/01/2022
Accepted: 13/02/2022
Published: 15/06 /2022

The blood components of a liquid cytoplasm have these kinds: RBCs, WBCs, and platelets. Acute and chronic leukemia are the two types of leukemia described by hematologists. There are two kinds of acute leukemia: acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) (Al-jaboriy et al., 2019). Acute type leads to be fatal within two or three months. The patients with a high L1 + L1/L2 ratio were significantly more Survival than that patients with L2/L1 + L2 ratio (Miller et al., 1981).

The work of (Sarrafzadeh et al., 2015) focused on the segmentation of white blood cells WBCs by K-means clustering blood smear microscopic images after image converted to Lab with proper Initial Seed Points (ISP) to segment nuclei. Other researchers (Ghane et al., 2017) introduced a method combined between k-means clustering, thresholding, and also used watershed algorithms to extract WBCs from microscopic images. the nuclei were extracted from cells and segmentation

was performed for overlapping cells and nuclei. (Hegde et al., 2019) suggested that a combination of mathematically operation and morphological processes combined to segment nuclei even in the presence of illumination changes by active contour, reporting a 96 % dice score. The active contours method was used for the detection of leukocytes considering nuclei as masks. (Ashour et al., 2021) proposed a Histogram-based Object to Background Disparity metric (HOBD) determined utilizing the green component histogram based features that were resulting from the at the start extracted WBCs, the result for dice score 92%. (Kadry et al., 2021) proposed various CNN-based segmentation schemes for the WBC Segmentation. The result for the dice score was 94.4%. (Tavakoli et al., 2021) suggested an algorithm for segmenting the WBC nucleus. Using an SVM classifier, three shapes and four color features are designed and segmented the cytoplasm. The result for the dice score was 96.75%. (Amin et al., 2015) to end of literature review proposed an approach to the detection of Acute lymphoblastic leukemia (ALL) and classified it into L1, L2, and L3 ALL subtypes. they applied the K- means algorithm for segmentation nuclei in cells. So the identification of the ratio of ALL subtypes is very important which can be done through image segmentation of those types. There were little works on Subtype ALL segmentation with their important determining survival rate. The aim of this work is to best segment the subtypes of ALL automatically and for different Dataset locally and publically.

2. MATERIALS AND METHODS

2.1. Material

Totally, 427 microscopic blood smear images were collected for testing the proposed system. The 55 images were gathered (ALL-IDB2). All of the images in the ALL-IDB2 are unhealthy cases. These images have a resolution of 257*257 pixels and a color depth of 24 bits (Labati et al., 2011).

Also, locally datasets compose 372 images, 186 images of them with a yellow filter and others without a yellow filter obtained from Nanakaly Hospital – Erbil. These images have a resolution of 3488 * 2616 pixels. Samples were obtained from peripheral blood and bone marrow for five distinct patients'. Leishman staining used for all of the slides in the locally databases. Image Acquisition was performed by using a digital light VanGrand microscopy supported by a digital camera (9 MP). The codes proceeded with MATLAB 2020a.

2.2. Methods

The method proposed in this work is divided basically into four steps, including image acquisition, preprocessing, segmentation, and post-processing, illustration in Figure 1.

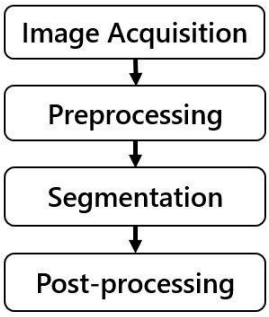

Fig. 1: describes the proposed algorithm's main steps.

2.2.1. Image Acquisition

All of the images (i.e. public and local) were confirmed by an experienced hematologist (works at Nanakaly Hospital-Kurdistan/Iraq) for ALL subtypes. Locally images were captured under the nearly same conditions, (i.e. lighting, position) for both yellow filter and without filter, as shown in Figure 2 (row B and C).

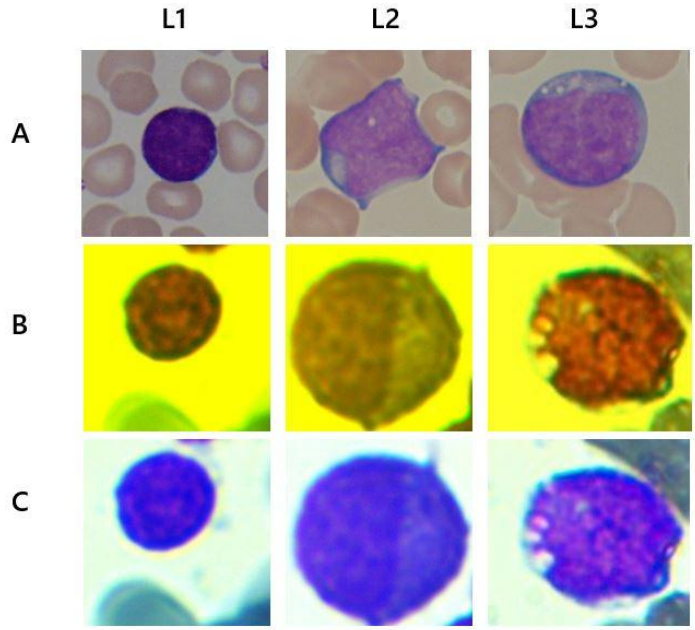

Fig. 2: row A shows a sample of Public dataset ALL_IDB2 (Shafique and Tehsin, 2018) (Shafique and Tehsin, 2018) and row B and C are for Local dataset with yellow filter and without yellow filter, respectively (for all subtypes of ALL) .

2.2.2. Preprocessing

One of the most important steps of preprocessing was to convert color space from one to another type. All color space models are described by hue (color shade), saturation (amount of gray or pure color), and luminance to describe colors (intensity, or overall brightness) (Tsai and Lee, 2002). Color space conversion is applied on blood smears to enhance their images. The HSV, XYZ, Lab, YCbCr, and YIQ color spaces are five color spaces that are compared and better of them (according to their dice score) utilized in the segmentation process.

2.2.3 Image Segmentation

The K-means clustering is utilized in the segmentation step after applying color space conversion, for separate nucleus and cytoplasm in ALL subtypes. It is applied using three clusters options to segment the nucleus, cytoplasm, and background region. K-Means is a signal processing, image clustering, and data mining unsupervised machine-learning algorithm. A set of n observations is partitioned into K clusters using K-means clustering. The cluster with the closest mean is assigned to each observation with each cluster's mean value serving as a pattern. As a result, the data space is partitioned into Interpolation cells. Initialization, computation, and convergence are the three stages of the algorithm Bhimani et al., 2015 .

2.2.4 Post Processing

In Post Processing the cluster one ignored for useless contains information. Clusters two and three were for cytoplasm and nucleus segmenting, respectively. For increasing performance (i.e. number of correct segmenting with dice greater than 0.9 / total number of images) this work proposed if operations (Al Hamad, 2013). If the center of cluster two is white passed to morphological operation by using remove the small object and closing the image. However, if the image in the center of cluster two is not white, the algorithm moves on to cluster three. If cluster three has a white pixel in the center the algorism moves again to morphological operation. But, if cluster three is not white the algorithm converts the image to a negative type. And goes to the morphological operation. In the morphology operation, we found a mask to finally segment cell images with black background (see Figure 5 row F). The proposed algorithm is appeared in Figure 3.

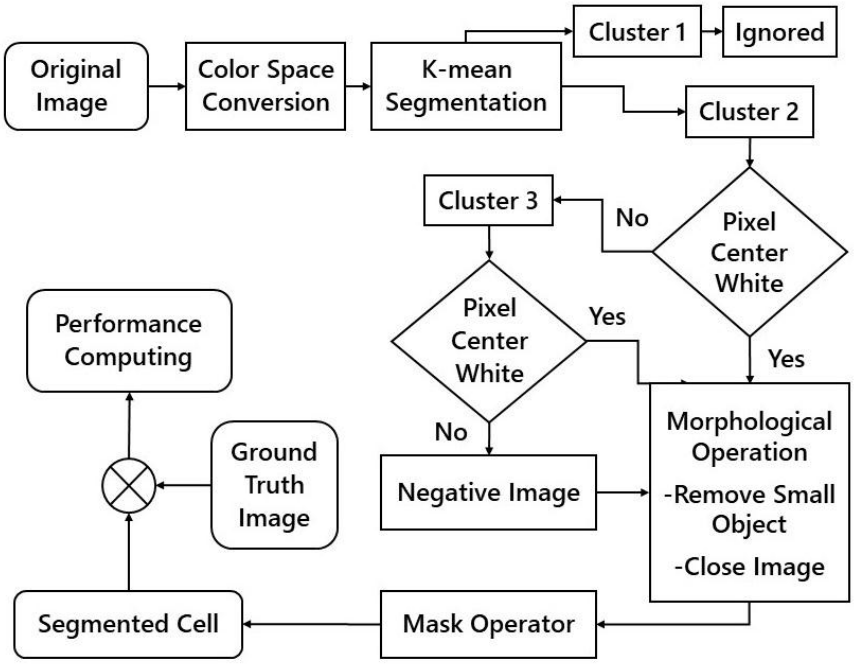

Fig. 3: Flowchart for proposed segmentation algorithm

In this work, five Segmentation Evaluation Parameters are used. There are necessary to computing segmentation qualities. all of the methods need ground truth image. The ground truth images were obtained through a code and confirmed by a hematologist (see the last row of figure 5)

Dice similarity coefficient (DSC): The Dice similarity coefficient represents spatial overlap, where DSC equals twice the number of elements common to intersected regions between Y, and X divided by the sum of region Y and region X. where the segmented image represents X and the ground truth represents Y. DSC is mathematically described as eq.1 (Zou et al., 2004):

$$\mathrm{DSC} = \frac{2|X \cap Y|}{|X| + |Y|} \qquad \ldots 1$$

Jaccard Similarity Coefficient (JSC): amounts the diversity and similarity of the finite sample sets calculated by the number of elements common to intersected regions between X and Y divided by union of region X and region Y as shown in eq.2 (Prabha and Kumar, 2016):

$$\mathrm{JSC} = \frac{|Y \cap X|}{|X \cup Y|} = \frac{|Y \cap X|}{|Y| + |X| - |Y \cap X|} \qquad \ldots 2$$

Probability Rand Index (PRI): During the data clustering process, the similarity between two regions is measured. It compares the segmented image to its ground truth for labeling consistency. It counts a certain number of pairs of pixels and averages the result across all of an image's ground truths. PRI is mathematically described as eq.3 (Zou et al., 2004).

$$R = \frac{A + B}{A + B + C + D} \qquad \ldots 3$$

Where A is the number of pairs of elements in the image that belong to the same subset in both regions X and Y. B, the number of pairs of image elements in distinct subsets in region X and different subsets in region Y. C, the number of pairs of image elements that belong to the same subset in region X but to separate subsets in region Y. D, the number of pairs of image elements in different subsets in region X but the same subset in region Y.

The distance between two clusters of mutual information is measured by the Variation of Information (VOI), which is a simple linear expression. Variation of Information computes the unpredictability in one segmentation in terms of distance from another segmentation, as shown in eq. (4) and Figure 4 (Prabha and Kumar, 2016)

$$\mathrm{VI}(Y, X) = \mathrm{H}(Y) + \mathrm{H}(X) - 2\mathrm{I}(Y, X) \qquad \ldots 4$$

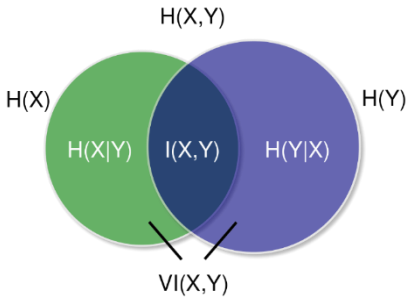

Fig. 4: Variation of Information parameters.

The Global Consistency Error (GCE) is an amount that determines how much one segmentation differs from another in terms of refinement. Related segmentations are decided consistent because they could represent the same image segmented at different scales. The term "segmentation" refers to the grouping of pixels in an image. Pixel groups make up the segments. The pixel is in a refinement zone if one segment is a perfect subset of the other, and the error should be 0. The two zones will overlap in an unforeseen way if there is no Subset relationship. eq.5 is the mathematical formula for GCE (Prabha and Kumar, 2016).

$$.\mathrm{GCE} = \frac{1}{n}\ min\left\{\sum_{i.} E(q_1, q_2, p_{i.}), \sum_{i.} E(q_1, q_2, p_{i.})\right\} \ldots\ 5$$

The segmentation error measure accepts two input segments, q1 and q2, and returns a real valued output in the range from zero to one, with 0 indicating no error. Consider the segments in q1 and q2 that contain that pixel.

3. RESULT AND DISCUSSION

Different color space types with their components were checked for high performance and accurate segmentation (that obtained segmenting quality parameters) of different datasets (locally and publically) (Aslan et al., 2011). The performance computes for different color spaces and different datasets (locally and publically). According to table 1, the best color space is the YIQ color space for Q-component due to their high-performance values than other color spaces. The performances were 87 %, 70 %, 54 %, and 65 % for the public datasets, locally dataset without filter, with filter, and all database, respectively.

Table 1: The effect of different color spaces with their components on the performance of proposed algorism for image segmenting.

Color Space	Public Dataset	Local Dataset		All Dataset
	55 images in ALL_IDB2	186 images without filter	186 images with filter	427 Images
Lab	80 % (a+b)*	65 % b	48 % L	57 % (a+b)
HSV	78 % S	44 % S	46 % (H+V)	43 % (H+S)
XYZ	58 % (X+Y)	38 % Y	42 % Y	42 % Y
YCbCr	76 % (Y+Cb)	54 % Cb	49 % Y	52 % Cb
YIQ	87 % Q	70 % Q	54 % Q	65 % Q

* inside parentheses represent component of color spaces

Figure 5 shows a sample of image segmenting steps for the image database. The first row (A) is for original images, the second row (B) is for YIQ color space conversion, the third row (C) is for segmented cytoplasm. The row (D-E) is the segmented nucleus, removed small objects and close operation, and finally segmented cell, respectively.

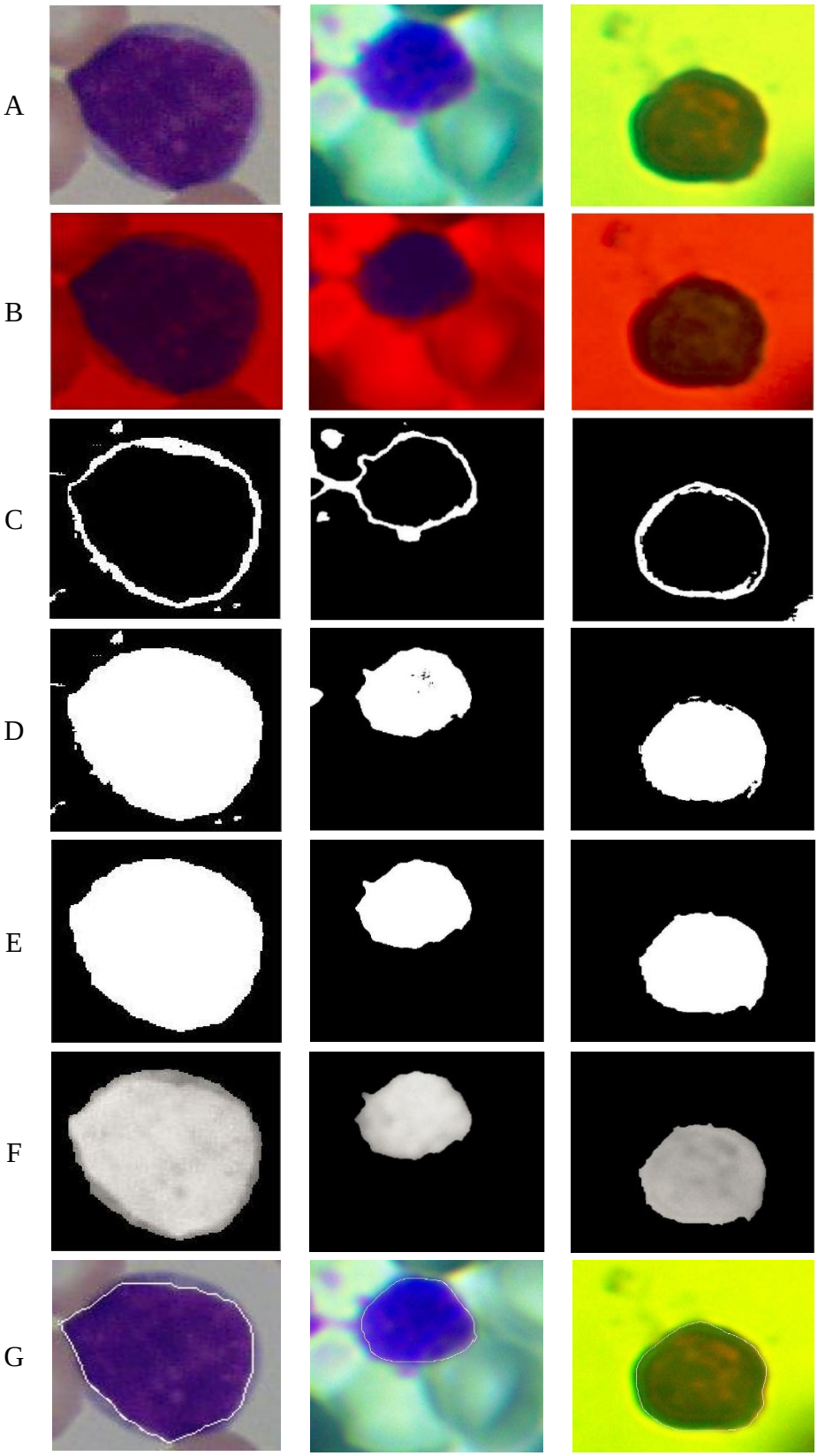

Fig. 5: show Original Cells (row A), YIQ color space (row B), cytoplasm segmented (row C), nuclues segmented (row D), removed small objects and close operation (row E), segmented cell (row F), and finally ground truth image (row G).

The proposed code can segment cytoplasm (row C of figure 5) and nucleus (row D) through results of clusters two and three of K-mean segmentation. One can see small unwanted objected and opening region in row D. By using morphological operations, there defected removed (see row E of the same figure). From the last results, a mask is obtained to finally segment the cells from the background see row E. The last row in the figure represents ground truth images for each database.

Corrected segmentation image (Dice > 0.9) for the number of different color spaces, and the best component displayed in Figure 6. The best color space for image segmentation for all datasets was YIQ color space it gave 48/55(%), 130/186(%), 141/186(%), and 279/427(%), corrected image per total number of images for public, local, and all datasets. Respectively. But the worst one was the XYZ components.

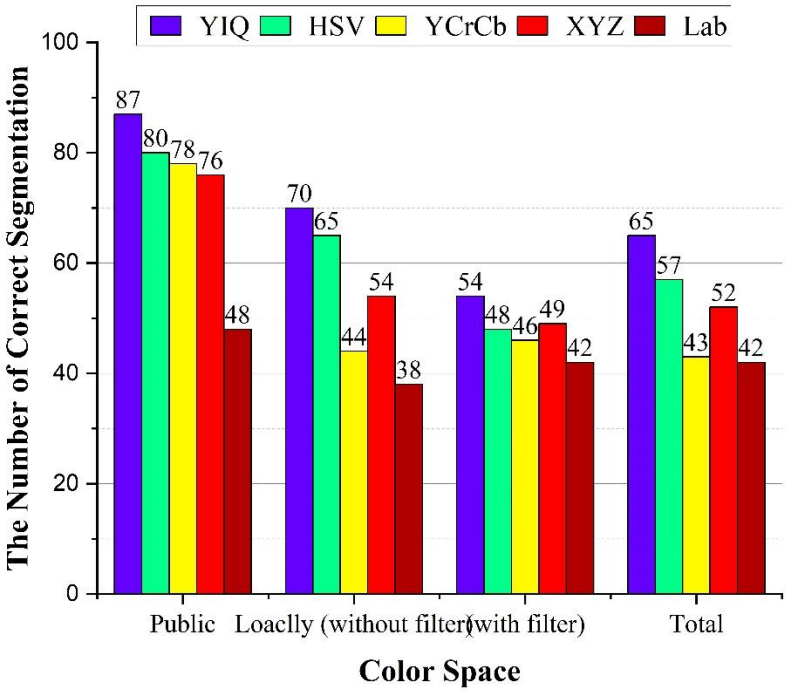

Fig. 6: The number of corrected segmentation images versus Color Space components for all databases.

For checking segmentation evaluation five parameters were used: Global Consistency Error (GCE), Dice Coefficient (DC), Jaccard Similarity Coefficient (JSC), Variation of Information (VOI), and Probabilistic Rand Index (PRI). Figure 7A shows the highest values of DC (0.90), JSC (0.93), and PRI (0.91) and the lowest values of GCE (0.07), and VOI (0.4) for NTSC color space which means best segmentation results for ALL subtypes. Also, other sub-images of figure 7 (i.e., B-D) show the highest values for DC, JSC, and PRI and the lowest values of GCE, and VOI for the same color space. So, the best color space segmentation results were from YIQ color space with Q component for all databases.

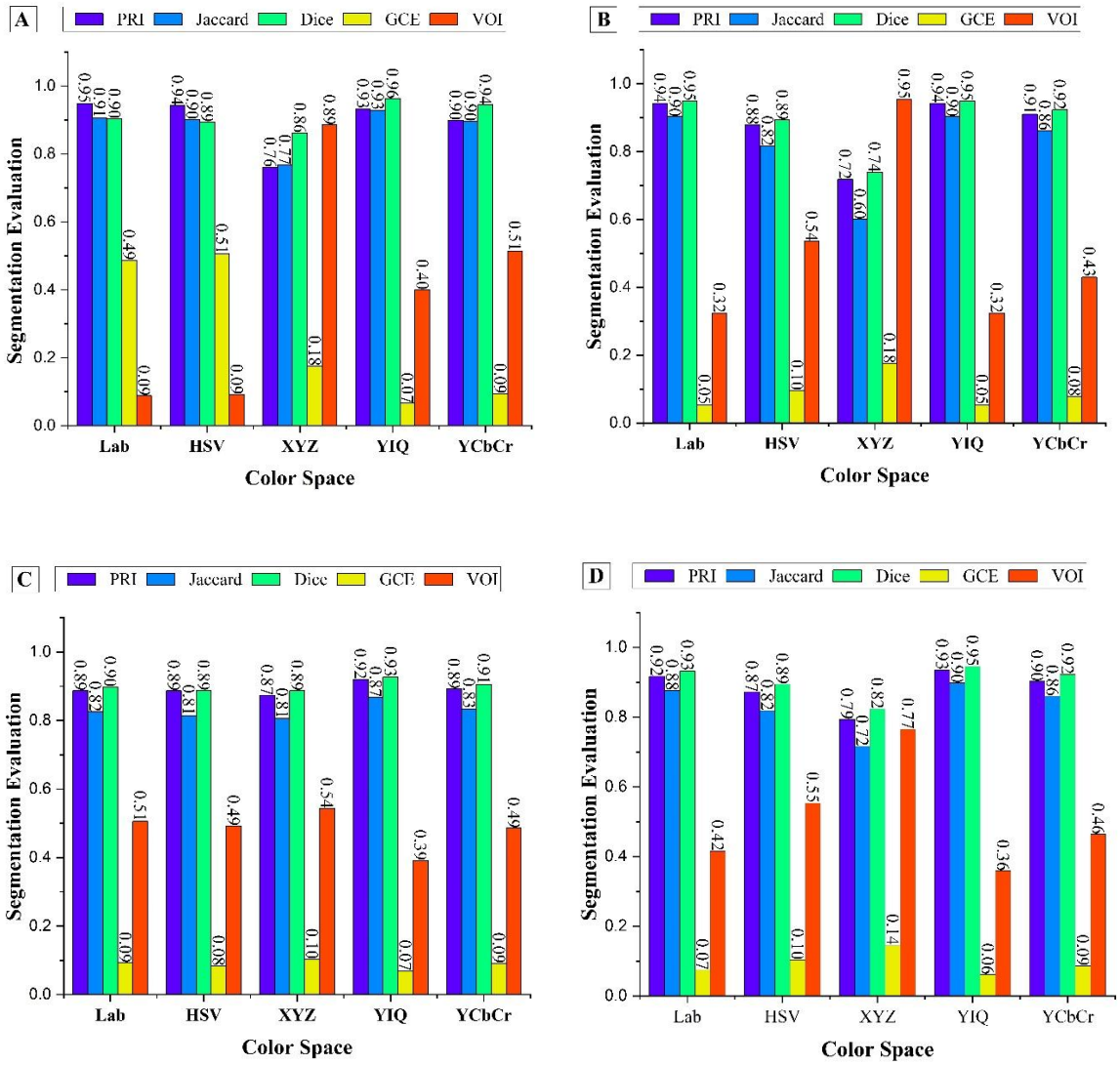

Fig. 7: display different segmentation evaluation parameters for several color spaces by using: public dataset (A), Local Dataset without Filter (B), Local Dataset with Filter(C), all dataset (D).

Table 2: comparing dice score values for the proposed method with other results of mentioned researchers for segmentation evaluation

Author, Year.	**Ashour et al., 2021**	**Kadry et al., 2021**	**Hegde et al., 2019**	**Tavakoli et al., 2021**	**Proposed Method**
Dice Score	92%	94.40%	96 %	96.75 %	96.24%

The compression result of the proposed algorism with other works is displayed in table 2. The research performed by (Ashour et al., 2021) obtained a Dice score of 92% for segmenting of basophila and esonophile. But they use segmentation based histogram and classification methods that take a lot of time to run. Also, the authors (Kadry et al., 2021) used classification shems to segment WBC components. They use different shems such as Segnet, U-net, and VGG-Unit. The better result was for VGG-Unit with a Dice score of 94.4%. the proposed method obtained the highest dice index (96.24%) with a very short time to run (31.27±22.8 ms) for automated ALL segmentations. Another research, (Tavakoli et al., 2021) segmented normal subtypes of WBC dice score 97.75%. but proposed work

targets auto-segmenting subtypes of ALL. The authors (Hegde et al., 2019) obtained 96 % of dice scores by using an algorithm that has large steps of image processing that needs optimization of a large number of parameters. So, in compassion with mentioned works, the proposed work is characterized by high performance, small running time, high quality segmenting, and finally automated algorism.

4. CONCLUSION

In this work, a new technique of the automatically segmentation of blast cells from microscopic blood smear images provided. This study presents a new combination of image processing methods and proposes significant pre-processing to gain excessive segmentation performance. In particular, the five designated color spaces with K-means cluster segmentation and the ability of morphological operations to differentiate the three subtypes of ALL was demonstrated. The selection of color space with their components based on their similarity with ground truth image through using five evaluation parameters. The proposed codes for ALL subtype accurate segmentation applied on local and public image datasets (427 images). the best color space type was YIQ which had 87% performance for the public dataset with segmentation evaluation 96.24% for dice parameter.

Acknowledgements

Thank you to the hematologists Dr. Hewa Ahmed, in the laboratory at the cancer cell Department of Nanakaly Hospital in Erbil, and the haematologists Dr. Hewa A. Mustafa in Rapert Teaching Hospital, for accepting me during my work. Thanks to the radiation oncologist, Dr. Beston S. Hassan Head of the Cancer Unit of the Health Department at Awat Center for Cancer Statistics in Erbil, Ministry of Health, Kurdistan Regional Government.

Conflict of Interest: The authors declared that they have no conflicts of interest.

References

AL-JABORIY, S. S., SJARIF, N. N. A., CHUPRAT, S. & ABDUALLAH, W. M. 2019. Acute lymphoblastic leukemia segmentation using local pixel information. *Pattern Recognition Letters,* 125**,** 85-90.

AL HAMAD, H. A. Use an efficient neural network to improve the Arabic handwriting recognition. 2013 IEEE International Conference on Signal and Image Processing Applications, 2013. IEEE, 269-274.

AMIN, M. M., KERMANI, S., TALEBI, A. & OGHLI, M. G. 2015. Recognition of acute lymphoblastic leukemia cells in microscopic images using k-means clustering and support vector machine classifier. *Journal of medical signals and sensors,* 5**,** 49.

ASHOUR, A. S., WAHBA, M. A. & GHANNAM, R. 2021. A Cascaded Classification-Segmentation Reversible System for Computer-aided Detection and Cells Counting in Microscopic Peripheral Blood Smear Basophils and Eosinophils Images. *IEEE Access*.

ASLAN, M. S., MOSTAFA, E., ABDELMUNIM, H., SHALABY, A., FARAG, A. A. & ARNOLD, B. A novel probabilistic simultaneous segmentation and registration using level set. 2011 18th IEEE International Conference on Image Processing, 2011. IEEE, 2161-2164.

BHIMANI, J., LEESER, M. & MI, N. Accelerating K-Means clustering with parallel implementations and GPU computing. 2015 IEEE High Performance Extreme Computing Conference (HPEC), 2015. IEEE, 1-6.

GHANE, N., VARD, A., TALEBI, A. & NEMATOLLAHY, P. 2017. Segmentation of white blood cells from microscopic images using a novel combination of K-means clustering and modified watershed algorithm. *Journal of medical signals and sensors,* 7**,** 92.

HEGDE, R. B., PRASAD, K., HEBBAR, H. & SINGH, B. M. K. 2019. Image processing approach for detection of leukocytes in peripheral blood smears. *Journal of medical systems,* 43**,** 1-11.

KADRY, S., RAJINIKANTH, V., TANIAR, D., DAMAŠEVIČIUS, R. & VALENCIA, X. P. B. 2021. Automated segmentation of leukocyte from hematological images—a study using various CNN schemes. *The Journal of Supercomputing***,** 1-21.

KARWAN, M., ABDULLAH, O., AMIN, A., HASAN, B., MOHAMED, Z., SULAIMAN, L., SHEKHA, M., NAJMULDEEN, H., BARZINGI, B. & SALIH, A. 2021. Cancer Statistics in Kurdistan Region of Iraq: A Tale of Two Cities.

LABATI, R. D., PIURI, V. & SCOTTI, F. 2011. The Acute Lymphoblastic Leukemia Image Database for Image Processing. *Universita Degli Studi Di Milano,* 10.

MILLER, D. R., LEIKIN, S., ALBO, V., SATHER, H. & HAMMOND, D. 1981. Prognostic importance of morphology (FAB classification) in childhood acute lymphoblastic leukaemia (ALL). *British Journal of Haematology,* 48**,** 199-206.

PHILIP, A. T., SHIFAANA, S., SUNNY, S. & MANIMEGALAI, P. Detection of Acute Lymphoblastic Leukemia in Microscopic images using Image Processing Techniques. Journal of Physics: Conference Series, 2021. IOP Publishing, 012022.

PRABHA, D. S. & KUMAR, J. S. 2016. Performance evaluation of image segmentation using objective methods. *Indian J. Sci. Technol,* 9**,** 1-8.

SARRAFZADEH, O., DEHNAVI, A. M., RABBANI, H. & TALEBI, A. A simple and accurate method for white blood cells segmentation using K-means algorithm. 2015 IEEE Workshop on Signal Processing Systems (SiPS), 2015. IEEE, 1-6.

SHAFIQUE, S. & TEHSIN, S. 2018. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technology in cancer research & treatment,* 17**,** 1533033818802789.

SIEGEL, R. L., MILLER, K. D., GODING SAUER, A., FEDEWA, S. A., BUTTERLY, L. F., ANDERSON, J. C., CERCEK, A., SMITH, R. A. & JEMAL, A. 2020. Colorectal cancer statistics, 2020. *CA: a cancer journal for clinicians,* 70**,** 145-164.

TAVAKOLI, E., GHAFFARI, A., KOUZEHKANAN, Z. M. & HOSSEINI, R. 2021. New Segmentation and Feature Extraction Algorithm for Classification of White Blood Cells in Peripheral Smear Images. *bioRxiv*.

TSAI, C.-M. & LEE, H.-J. 2002. Binarization of color document images via luminance and saturation color features. *IEEE Transactions on Image Processing,* 11**,** 434-451.

ZOU, K. H., WARFIELD, S. K., BHARATHA, A., TEMPANY, C. M., KAUS, M. R., HAKER, S. J., WELLS III, W. M., JOLESZ, F. A. & KIKINIS, R. 2004. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology,* 11**,** 178-189.