Al-Nahrain Journal for Engineering Sciences **NJES** 26(4)337-342,2024
http://doi.org/10.29194/NJES.26040337

Detection of Megakaryocyte Cell Structure Through Artificial Intelligence Tools

Shaima Ibraheem Jabbar[1]and Abathar Qahtan Aladi[2]

Authors affiliations:
1) Medical Instruments Technologies, Al-Furat Alawsat Technical University, Babylon Technical Institute-Iraq.
shaima.jabbar@atu.edu.iq

2) Babylon Health Directory, Mirjan Teaching Hospital
aladiabathar@gmail.com

Paper History:

Received: 18th Aug. 2023

Revised: 3rd Sep. 2023

Accepted: 11th Sep. 2023

Abstract

Recent research has focused on analysing megakaryocyte images to extract the information needed to track the progression of nervous system diseases. Segmentation is a fundamental step in describing and analysing the core contents of megakaryocytes, including the cytoplasm and nucleus. In this study, 45 megakaryocyte images were obtained. A new segmentation image technique was proposed, called the updating fuzzy c-means technique, through the intelligent selection of the centres of each cluster to separate cell components. The first step of this technique (fuzzification) was based on a knowledge analysis of the local parameters (entropy, contrast and standard deviation) that had a substantial influence on the grey-level distribution between the cytoplasm and nucleus. The second important step was the construction of fuzzy rules in terms of the variation in these local parameters to control the intelligent pick-out or update the centroid of each cluster and obtain a successful separation of the cytoplasm and nucleus. The final step was defuzzification to obtain the output images. The results revealed the superiority of the proposed method over recent technique. The accuracy of the segmented nucleus was greater than 7.46%; in the case of the cytoplasm, the accuracy was higher at 18%. These results indicated that this technique may be applied on other biomedical images.

Keywords: Megakaryocyte Images, Artificial Intelligence, Fuzzy C-Means Technique, Fuzzy Inference Technique.

الكشف عن مكونات الخلايا الجذعية باستخدام أدوات الذكاء الاصطناعي

شيماء ابراهيم جبار، اباذر قحطان العادي

الخلاصة:

ركزت الأبحاث الحديثة على تحليل صور الخلايا الجذعية لاستخراج المعلومات الواصفة لمكوناتها والضرورية لتتبع تطور الأمراض العصبية. ان فصل مكونات الخلايا الجذعية هو احدى الخطوات الأساسية في وصف وتحليل المحتويات الأساسية لهذه الخلايا، بما في ذلك السيتوبلازم والنواة. في هذه الدراسة، تم الحصول على ٤٥ صورة للخلايا الجذعية. تم اقتراح وتطبيق تقنية صور تجزئة جديدة، تسمى تقنية تحديث الوسائل الضبابية. ان الخطوة الأولى من هذه التقنية هي التضبيب والذي يعتمد على تحليل العوامل التي توصف طبيعة الصورة وخواصها (الانتروبيا والتباين والانحراف المعياري)،ان هذه العوامل بطبيعتها لها تأثير كبير على توزيع المستوى الرمادي بين السيتوبلازم والنواة. اما الخطوة الثانية هي بناء قواعد غامضة تلخص تأثير هذه العوامل على الانتقاء الذكي أو تحديث النقطة الوسطى لكل عنقود(السينترويد) لفصل السيتوبلازم عن النواة بشكل ناجح. ان الخطوة الأخيرة هي إزالة االتضبيب للحصول على الصور الناتجة. أظهرت النتائج تفوق التقنية المقترحة في هذا البحث نسبة الى تقنية حديثة منفذة. حيث كانت دقة النواة المجزأة أكبر من ٧,٤٦٪. ، وكانت الدقة أعلى بنسبة ١٨٪ في حالة السيتوبلازم. وكذلك أشارت هذه النتائج إلى أنه بالإمكان تطبيق هذه التقنية على صور طبية حيوية أخرى.

NJES is an **open access Journal** with **ISSN 2521-9154** and **eISSN 2521-9162**

1. Introduction

Automated morphological feature extraction of megakaryocytes is a key pathological method for the detection of bone marrow diseases. These diseases affect the morphological structure of megakaryocytes [1–5]. The delineation and isolation of the core content of a cell are fundamental steps in the extraction of its morphological properties [6,7]. Figure (1) illustrates the manual labelling of the main structure of megakaryocytes.

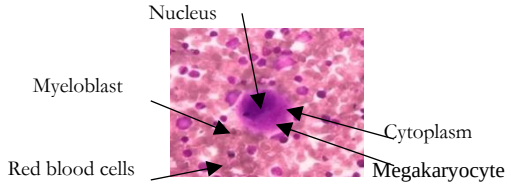


Figure (1): Shows labeling of megakaryocyte components

From the figure above, the main challenge faced by many researchers is the classification of grey-level intensities, which belong to the megakaryocyte cell components (nucleus and cytoplasm) [8–11]; this task remains strenuous. The main reason for this is the texture and colour similarity of these components.

Several studies have been published on medical image analysis [12,13] and megakaryocyte image segmentation [14]. These studies have used traditional techniques and artificial intelligence tools to obtain separate nuclear and cytoplasmic images. Fuzzy clustering has been used to isolate nuclear structures in biopsy images [15]. However, the selection and updating of the cluster center remain the most important challenges in this technique. Deep learning is a successful tool that has been used to detect the boundaries of megakaryocyte structure [16,17]. This method is particularly useful for extracting considerable features, such as edges, from an image. This technique helps consultants obtain a high-level stage of diagnosis. Thresholding of the coloured leukemic blast nucleus was performed using a traditional method (Otsu's technique and morphological operations) [18]. However, these techniques are still classical for isolating atypical shapes of gait structures. In the modified fuzzy c-means technique, an appropriate centroid can be selected by intelligent detecting the closest Euclidean distance between each pixel of the image and centroid based on the fuzzy inference technique. This technique was applied to megakaryocyte cells [19] is called modified fuzzy technique one (MFCT1) but only on a small set of data. Centroid detection was insufficient, depending on the Euclidean distance. Hence, in this study, we proposed and applied a new technique based on the study of the effect of differences in local image properties (fuzzy entropy [FN], standard deviation [SD] and contrast [C]) to detect suitable centroids. Megakaryocyte images are colour images that must be disintegrated into their original components (red, green and blue) before using the proposed technique. On the basis of the colour of the input image in Figure (1), the dominant component was red because megakaryocyte images under a microscope tend to be red. In the next stage, megakaryocyte image data were mapped from the grey-level intensity scale to a fuzzy set scale through a membership function. FN, SD and C were evaluated as metrics to detect the variability of grey-level intensities, which resulted from overlapping grey-level intensities for cell structures (megakaryocyte images). Smart detection of grey-level intensities helped reach the active centroid, which is an important step in image segmentation.

The remainder of this paper is organized as follows. The background illustrates the main components of the proposed technique in section one. Section two presents the structure of the remodeled fuzzy c-means technique. The experiments and results are presented in section three. In section four, we discuss the results. Section five concludes the paper.

1.1 Background

Different techniques are formally defined in this section as follows.

1.1.1 Local image features

The local image features include the evaluation of SD, C and FN of an image. Homogenous regions had a low level of fuzzy entropy, whereas regions with greater detail had a high level of fuzzy entropy. Therefore, a high fuzzy entropy value is an indicator of selecting a suitable centroid. The extraction of fuzzy entropy (FN) depends on the determination of Shannon's entropy [20,21]: The equation of Shannon's entropy is:

$$N\left(\mu_T(F_{xy})\right) = \frac{1}{r_1 * r_2 * \ln(2)} X\left(\sum_y^Y \sum_x^X T(\mu_T(F_{xy}))\right) \quad \ldots. (1)$$

Where $r_1$and r_2 represent the dimensions of image, F_{xy} is input sub-image (3x3), μ_T is membership function and T(.) is Shannon's entropy function, given by:

$$T\left(\mu_T(F_{xy})\right) = -\left(\mu_T(F_{xy})\right) \times \ln\left(\mu_T(F_{xy}) - \mu_T(F_{xy}\right) \times \ln\left(1 - \mu_T(F_{xy})\right) \quad \ldots. (2)$$

1.2 Previous Technique

In previous techniques, fuzzy c-means was used to classify clusters [22,23]. The number of clusters can be determined based on the identification of the image objects that are separated. Grouping data samples together (clusters) is based on measuring the minimum distance between each pixel and centroid of the cluster. The following equations describe the evaluation of this measurement. This equation is defined as follows: m is greater than 1, g_{ij} is the fuzzy mapping function of px_i in cluster j, px_i is the ith dimension of the measured data (image pixels) and $cent_j$ is the centroid. Equations (4) and (5) were utilized to update and select the cluster center $cent_j$ using Equations (3) and (4):

$$H_m = \sum_{i=1}^{k} \sum_{j=1}^{cent} g_{ij}^m \left\| px_i - cent_j \right\|^2 \quad 1 \le m < \infty \quad \ldots\ldots (3)$$

The definition details of this equation as follows: the value of m is greater than 1, g_{ij} is fuzzy mapping function of px_i in the cluster j, px_i is the ith of dimensional measured data (image pixels) and $cent_j$ is a centroid. Equation (4) and equation (5) have been utilized to update and select the cluster centers $cent_j$ using equation (3) and equation (4):

$$g_{ij}=\frac{1}{\sum_{r=1}^{cent}\left(\frac{\|px_i-cent_j\|}{\|px_i-cent_r\|}\right)^{\frac{2}{m-1}}} \quad \ldots.(4)$$

The center of cluster (centroid) represents the mean of all points, weighted by their membership of belonging to the cluster; therefore, this method is called fuzzy c-means.

$$cent_j=\frac{\sum_{i=1}^{k} g_{ij}^{m}.px_i}{\sum_{i=1}^{k} g_{ij}^{m}} \quad \ldots..(5)$$

The center of the cluster (centroid) represents the mean of all points weighted by their membership in the cluster; therefore, this method is called fuzzy c-means. The process stops when converging to a local minimum Hm. Recently, image components have been isolated via a modified fuzzy c-means method [19] that employs a fuzzy inference system based on the measurement of Euclidean distance. However, distance is not always considered a sufficient factor for base selection. Therefore, other parameters that affect the chromaticity of the image may be relied upon. In this study, we updated the method by relying on parameters (FN, C and SD) and without performing any pre-processing on the input image, such as wavelet processing.

2 Materials and Methods

The center of the cluster (centroid) represents the mean of all points weighted by their membership in the cluster; therefore, this method is called fuzzy c-means. The process stops when converging to a local minimum Hm. Recently, image components have been isolated via a modified fuzzy c-means method [19] that employs a fuzzy inference system based on the measurement of Euclidean distance. However, distance is not always considered a sufficient factor for base selection. Therefore, other parameters that affect the chromaticity of the image may be relied upon. In this study, we updated the method by relying on parameters (FN, C and SD) and without performing any pre-processing on the input image, such as wavelet processing.

2.1 Protocol of Data Collection

In this study we collect about 45 photographic image samples from about 15 patients (7 males and 8 females), age range (6-58) years. The images of 10 patients are obtained from [19]. The permission was acquired from those patients for data collection and consent letters were signed from patients in accordance with institutional ethics guidelines (Ref 4506MC). The bone marrow aspirate was requested for different causes, these samples show normal bone marrow cells, at first the marrow sample are aspirated from posterior iliac fossa of the patient, putted in to EDTA tubes, then spread on glass slides, let it dried and then stained by specific stain which was geimsa stain. The next step was putting few drops of this stain on dried slides and mixed with few drops of water, waiting about 5 minutes and then washed with water. After that, lets dried again and then wiped with oil for proper visualization of cells under microscope. The nucleus appeared deep purple color and cytoplasm were pinkish in color. The examination done under power 40X of Olympus microscope. 512 x 512 pixels is a dimensions of acquired image, these images are color images (R, G,B), each component has 8 bits of intensity resolution.

2.2 Modified Fuzzy c-means technique (MFCT2)

The proposal technique consists from a set of steps, the following flowchart illustrates of proposed technique in the figure below:

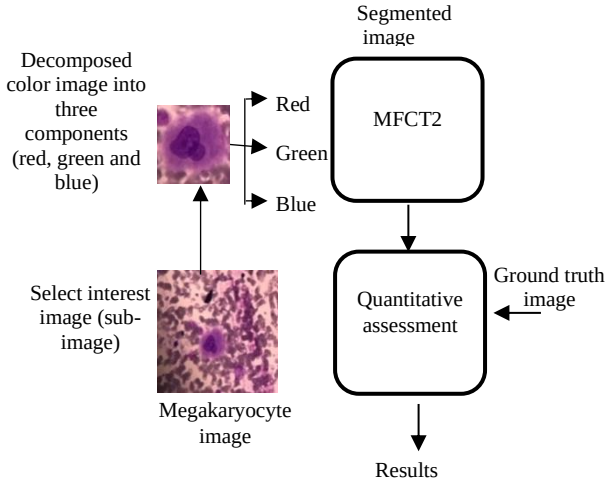


Figure (2): shows flow chart of segmentation procedure and quantitative assessment results.

In the fuzzy c-means technique, Euclidean distance is a useful attribute for converging to the solution and labelling image pixels into a particular cluster. However, it is sometimes insufficient to define all the points belonging to this cluster. Some points had a small distance and were classified within this cluster. However, they did not belong to this cluster because of intersecting image pixels. Therefore, in this study, the fuzzy c-means technique was updated to include all points. Image information can be classified, and clusters depend on the analysis of image features such as grey-value intensities and the measure of the fuzziness metric. It can update the center of the cluster and the membership degree based on these features instead of only optimizing the distance between the image data and the center of the cluster. The center of the cluster and membership degree were updated using Equations (3) and (4). Therefore, we conscripted eight fuzzy rules as follows (**En** is entropy, **SD** is the standard deviation and **C** is contrast):

Rule 1: If **FE** is high, **SD** is high and **C** is low then updates centroid.
Rule 2: If **FE** is high, **CD** is low and **C** is low then pick out centroid.
Rule 3: If **FE** is high, **CD** is high and **C** is low then update centroid.
Rule 4: If **FE** is low, **CD** is low and **C** is high then update centroid.
Rule 5: If **FE** is low, **CD** is high and **C** e is high then update centroid.
Rule 6: If **FE** is low, **SD** is low and **C** is low then pick out centroid.
Rule 7: If **FE** is high, **CD** is high and **C** is high then update centroid.

Rule 8: If **FE** is low, **CD** is high and **C** is low then pick out centroid.

The algorithm steps of the modified fuzzy c-means technique are summarized as follows:	
Algorithm	
Step 1	*Input colour megakaryocyte images, decompose colour image (RBG) into three components (Red, Green and Blue) then select the most suitable component. From this image select the interest area of the image (sub-image).*
Step 2	*Assign the number of clusters (objects) and vector of image features will be involved in the fuzzy c-means algorithm such as standard deviation (SD), fuzzy entropy (FN) and Contrast (C). This depends on the objective of segmentation, in the case of megakaryocyte cell, two objects (nucleus and cytoplasm) will be extracted.*
Step 3	*Fuzzification using membership matrix G=[g_{ij}] and evaluate standard deviation (SD), fuzzy entropy (FN) and Contrast (C).*
Step 4	*Start with first value of the center of clusters and compute the distance of each data point to the centers using equation (4) and equation (5) respectively.*
Step 5	*Check standard deviation (SD), fuzzy entropy (FN) and Contrast (C) of each point of the image based on the eight fuzzy rules to make a decision either renew the fuzzy partition matrix and clustering center or not.*
Step 6	*Determine membership degree of each point of image dataset to each center.*
Step 7	*Based on the calculated membership degree compute the new centers of each cluster using equation (4) then go to the step 5*
Step 8	*Repeat steps until achieve the condition in the fuzzy rules and pick out convenient center and membership degree of pints of cluster. Detect the similarity between unlabeled data and centroid in term of the values of standard deviation (SD), fuzzy entropy (FN) and Contrast (C).*
	End

3. Experiments and results

After the subsequent steps, the first colour megakaryocyte image was decomposed into three components (red, green and blue). The red component was selected because the colour of the image tended toward this colour, and the image was clearer than the other components (see the figure below):

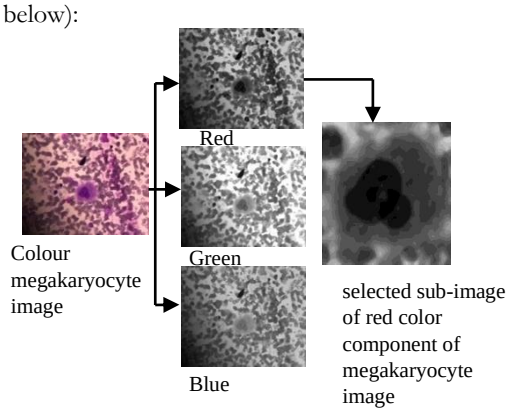


Figure (3): represents the color image acquired (RBG) and the one obtained after decomposition

The updated fuzzy c-means clustering method was applied to megakaryocyte images after pre-processing in the previous steps. The proposed image segmentation technique was compared with another image segmentation method described in a previous study [19]. Figure (4) illustrates the output images obtained using these techniques.

The next stage was to verify the validity of the computed measurements in all growth phases using quantitative metrics (sensitivity, specificity, recall, precision and accuracy), which required the estimation of the agreement between the segmented images and ground truth images (reference images). This was achieved by calculating a set of image information: true positive, true negative, false negative and true negative [24,25]. These metrics were applied to evaluate the performance of both methods for the segmentation of the nucleus and cytoplasm based on the obtained cluster. Thus, the cytoplasm was segmented into a particular cluster, and the nucleus had another cluster. Efficacious techniques can be developed for segmenting nuclei and cytoplasm.

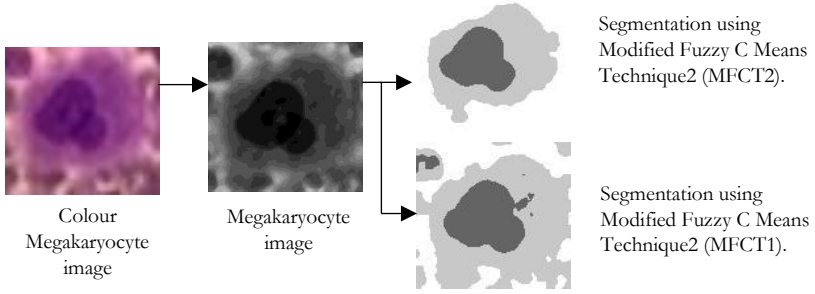
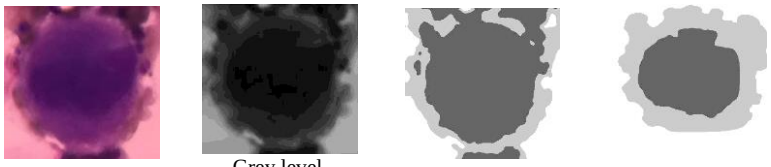


Figure (4): shows output images after applying MFCT2 and MFCT2**.**



Color Megakaryocyte image | Grey level intensities of megakaryocyte image | Segmentation using MFCT1 | Segmentation using MFCT2

Figure (5): presents example 1 (sample 5)

As shown in Figure (5), sample 5 of the image data demonstrated the remarkable performance of the proposed technique compared with the result of image detection using MFCT1. The shapes of the nucleus and cytoplasm were clear and less noisy. The resultant image obtained using the fuzzy c-means segmentation technique included additional information about the nucleus and cytoplasm. This hindered the complete separation of the morphological features of the cytoplasm and nucleus in the megakaryocyte images. The extraction of image properties significantly supports centroid detection and is more accurate than the MFCT1, which depends only on the evaluation of the closest distance. For most image samples, the proposed method showed much better accuracy than MFCT1. Considering 45 samples of megakaryocyte images, the accuracy of cytoplasmic clustering was approximately 18% higher than that of the fuzzy c-means technique. The mean accuracy of the nucleus presented as a cluster using the proposed method was approximately 7% higher than that of the same cluster obtained using the MFCT1.

4. Discussions

The evaluation of the closest Euclidean distance between the image pixels and centroid is the major factor that controls clustering in the fuzzy c-means technique. However, the dominant limitation of this technique is that it specifies a suitable centroid for a cluster [26,27]. The Euclidean distance cannot provide indications for grouping pixels as clusters, particularly in medical images (grey-level intensities interrupted with each other). Pixels that are close to the centroid but belong to another region can be grouped together. Furthermore, recently, we carried out recent technique (MFCT1) that depends on the intelligent selection of Euclidean distance [19]. However, it is not sufficient to extract all pixels set which describe megakaryocyte components due to this image is affected by other factors. Therefore, we proposed and updated the modification of fuzzy c means technique to improve the performance of the megakaryocyte components separation (MFCT2). The quantitative evaluations (sensitivity, specificity, recall, precision and accuracy) showed the superiority of the recent modified fuzzy c-means compared with previous technique (MFCT1), as shown in Figure (5). The dependence on the extraction image properties supports accurate centroid detection.

6. References

[1] A. A. Fauser and H. A. Messner, "Identification of megakaryocytes, macrophages, and eosinophils in colonies of human bone marrow containing neutrophilic granulocytes and erythroblasts", Blood, 53(5), pp. 1023–1027,1979

[2] A. Gonzalez-Villalva, P. Bizarro Nevaraes, M. Rajas-Lemus, M. Ustarroz-Cano, N. Lopez-Valdez, I. Garcia-Pelaez , J. Carlos Albarran-Alonso, F. Barbosa-Barron and T. Fortoul, "A brief review of the biology of megakaryocytes and platelets and their role in thrombosis associated with particulate air pollutio", Toxicology and Industrial Heath, 37(3), 164-172, 2021.

[3] B. W. Smith and G. P. Murphy, "Stem cells, megakaryocytes, and platelets", Curr. Opin. Hematol, 21:430-437, 2014.

[4] L. M. Bush, C. P. Healy, J. E. Marvin . *et al.* (2021) High-throughput enrichment and isolation of megakaryocyte progenitor cells from the mouse bone marrow, *Sci Rep* **11,** 8268, 2021.https://doi.org/10.1038/s41598-021-87681-2

[5] B. S. Wilkins, W. N. Erber, D. Bareford, et al., "Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes". Blood, 111(1):60-70, 2008.

[6] F. Valafar, "Pattern Recognition Techniques in Microarray Data Analysis", Annals of the New York Academy of Sciences, 980 (1): 41–64, 2002.

[7] K.Ohshima, M. Kikuchi and M. Takeshita، "megakaryocyte analysis of the bone marrow in patients with myelodysplastic syndrome, myeloproliferative disorder and allied disorders", Journal of Pathology, 177 (2), 181–189, 1995.

[8] V. Franco and *et al.* "Megakaryocytic features useful for the diagnosis of myeloproliferative disorders can be obtained by a novel unsupervised software analysis", Histology and Histopathology, 21(8), pp. 813–821, 2006.

[9] A. Bouzid-Daho, M. Boughazi, C. Tanougast and O. Benali Medjahed, "Textural analysis of bio-images for aid in the detection of abnormal blood cells", Int. J. Biomedical Engineering and Technology, Vol. 25, No. 1, pp.1–13, 20, 2017.

[10] F. A. Choudry, F. O. Bagger, I. C. Macaulay, S. Farrow, F. Burden, C. Kempster, H. McKinney, L. R. Olsen, N. Huang N and *et al.* "Transcriptional characterization of human megakaryocyte polyploidization and lineage commitment". J Thromb Haemost. 2021 May;19(5):1236-1249. 2021, doi: 10.1111/jth.15271.. PMID: 33587817.

[11] R. Verma and *et al.*, "A Multi-organ Nuclei Segmentation and Classification Challenge",

in *IEEE* Transactions on Medical Imaging, 2020. doi: 10.1109/TMI.2021.3085712.

[12] S. I. Jabbar, H. Majdi and A. Aladi, "Techniques of musculoskeletal system imaging", International journal of online Biomedical Engineering, 7(2) 1-11, 2022.

[13] X. Liu, K. Gao and et al., "Advances in Deep Learning-Based Medical Image Analysis," Health Data Science,14, 2021.https://doi.org/10.34133/2021/8786793

[14] M. Zhao, H. Wang, Y. Han, X. Wang, H. Dai, X. Sun, J. Zhang, M. Pedersen, SEENS: Nuclei segmentation in Pap smear images with selective edge enhancement", Future Generation Computer Systems, 114, 185-194, 2021.

[15] K. Sirinukunwattana. and *et al.*, Artificial intelligence–based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients", Blood advances, 4(14),3284-3294, 2020.doi:10.1182/bloodadvances.

[16] V. Kate, & P. Shukla, "Breast Cancer Image Multi-Classification Using Random Patch Aggregation and Depth-Wise Convolution based Deep-Net Model". International Journal of Online and Biomedical Engineering (iJOE), 17(01), pp. 83–100, 2021. https://doi.org/10.3991/ijoe.v17i01.18513

[17] E. Shweikeh, J. Lu and M. Al-Rajab. "Detection of Cancer in Medical Images using Deep Learning". International Journal of Online and Biomedical Engineering (iJOE), 17(04), pp. 83–100. 2021, https://doi.org/10.3991/ijoe.v17i14.27349

[18] S. Shinde, N. Sharma, P. Bansod, M. Singh and C.K. Singh, "Automated Nucleus Segmentation of Leukemia Blast Cells : Color Spaces Study", 2nd International Conference on Data, Engineering and Applications (IDEA), pp. 1-5, 2020, doi: 10.1109/IDEA49133.2020.9170721.

[19] S. I. Jabbar, L. M. Al Sharifi, "Megakaryocyte Images Segmentation Using Wavelet Transform and Modified Fuzzy C-means Techniques". In: Tran, DT., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, TD. (eds) Intelligent Systems and Networks . ICISN 2021. Lecture Notes in Networks and Systems, vol 243. Springer, Singapore, 2021 https://doi.org/10.1007/978-981-16-2094-2_65

[20] A. D. Luca & TS. ermini, "A definition of non-probabilistic entropy in setting of fuzzy set theory", *Information Control*, 20, 301–312, 1971.

[21] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier", Expert Systems with Applications, 38, 4600-460, 2011.

[22] J. C. A. Dunn, "Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics, 3: 32-57, 1973.

[23] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.

[24] D. M. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", *Journal of Machine Learning Technologies*, 2 (1): 37–63, 2011.

[25] A. F. Aanish and *et al.* (2018) The Use of Sensitivity and Specificity in Verification of Hadith's Chain of Transmission Centuries Ago, Sudan, Med J, 54 (3): 170–175, 2018.

[26] S. Gupta, "Medical image registration based on fuzzy c- means clustering segmentation approach using SURF", *Int. J. Biomedical Engineering and Technology*, Vol. 20, No. 1, pp.33–50, 2016.

[27] F. Bunyak and *et al.* "Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets", in Software tools and algorithms for biological systems, pp. 413–424, 2011.