

A New Augmented K-Means Algorithm for Seed Segmentation in Microscopic Images of the Colon Cancer

Ulaş YURTSEVER, Hayrettin EVIRGEN, Mustafa Cihat AVUNDUK

Abstract: In this study, we analyze histologic human colon tissue images that we captured with a camera-mounted microscope. We propose the Augmented K-Means Clustering algorithm as a method of segmenting cell nuclei in such colon images. Then we compare the proposed algorithm to the weighted K-Means Clustering algorithm. As a result, we observe that the developed Augmented K-Means Clustering algorithm decreased the needed number of iterations and shortened the duration of the segmentation process. Moreover, the algorithm we propose appears more consistent in comparison to the weighted K-Means Clustering algorithm. We also assess the similarity of the segmented images to the original images, for which we used the Histogram-Based Similarity method. Our assessment indicates that the images segmented by the Augmented K-Means Clustering algorithm are more frequently similar to the original images than the images segmented by the Weighed K-Means Clustering algorithm.

Keywords: cancer detection; clustering algorithms; histopathological image analysis; image segmentation; k-means

1 INTRODUCTION

According to a report issued by the World Health Organization (WHO) in 2008, 12.4 million new cancer cases and 7.6 million deaths due to cancer occurred worldwide. The top three most common cancers were lung (1.52 millions of people), breast (1.29 millions of people) and colorectal (1.15 millions of people) cancers [1]. Since cancer incidence is currently on the rise, identification of cancer via pathological imaging is of great importance.

Pathology is a field of science that is concerned mainly with diagnosis. Pathologic diagnostic procedures focus primarily on macroscopic and microscopic features. Pathologists make effort to determine abnormal changes in cells, tissues and organs. In order to reach an accurate diagnosis, the pathologist must rely on its accumulated clinical experience and expertise [2]. Therefore, the diagnostic process is inherently non-formulaic and in some cases a subjective judgment call has to be made. As objectivity and reproducibility are among the hallmarks of empirical sciences, various image analysis methods are proposed to replace the subjective judgment of the expert with an objective procedure that relies on quantifiable data. One of the most prominent among these analysis methods is image segmentation.

Image segmentation is the process of dividing an image into meaningful sub-objects. It involves classification of pixels and similar features in an image into distinct clusters [3]. Image segmentation plays a significant role in medical and microscopic imaging. In particular, it is used in the quantification of tissue volumes, localization of seed-gland-cell, localization of pathology, diagnosis, study of anatomical structure, and computer-integrated surgery [4, 3]. Furthermore, image segmentation is practical for many other applications such as the identification of possible areas of interest inside an image, and data labelling [5].

There are a great number of algorithms in the image segmentation literature. Such algorithms are divided into 5 main groups in general [6-9], namely, region-based segmentation algorithms [10-14], edge-based segmentation algorithms [15, 16, 17], thresholding-based segmentation algorithms [18, 19], computational or

clustering-based segmentation algorithms [20-26] and graph-based segmentation algorithms [27, 28, 29]. This study proposes an improved version of K-Means Clustering algorithm which is a clustering-based segmentation algorithm.

In the said literature, we come across several previous studies using clustering-based segmentation algorithms. Albayrak [30, 31] for instance, modified the K-Means Clustering algorithm---a classification method dependent on distance in the RGB and CIE Lab colour spaces---with the aim of making it clinically practical. Albayrak's approach made use of histogram data, and led to the invention of a new algorithm that he called Weighted K-Means Clustering. This image enables the re-creation of a set of given images with the colour classes generated by the algorithm and the identification of the re-created images that are more similar to the originals. Using over 120 images, Albayrak was also able to conduct tests to compare his algorithm to the K-Means Clustering algorithm. His results indicate that the algorithm he developed was better than the traditional K-Means Algorithm in colour clustering processing.

Mignotte [32] developed a simple (k-means based) segmentation strategy based on a fusion procedure. The fusion procedure aims to combine several segmentation maps in order to generate more consistent and accurate segmentation results. The proposed algorithm was tested on the Berkeley segmentation database, using over 300 colour images. The performance criteria used in the comparison were the probabilistic rand index (PRI), the variation of information (VoI), the global consistency measure (GCE), and the boundary displacement error (BDE). The results of the test indicate that the proposed fusion method---though simple and fast---performs (on the Berkeley natural image database) competitively and often better than the state-of-the-art segmentation methods, in terms of visual evaluations and quantitative performance measures. The database also contains ground truth segmentations obtained from human subjects for the purposes of quantitative evaluation.

Khan and Ahmad [33] also propose an algorithm to compute initial cluster centres for K-Means Clustering. This procedure is based on the experimental fact that very

similar data points (patterns) form the core of clusters and their cluster membership remains the same. Cluster centre initialization algorithm (CCIA) generates clusters which may be more numerous than the desired number of clusters. Similar clusters are therefore merged using density-based multiscale data condensation method to obtain the desired number of clusters. Khan and Ahmad used fossil data, wine recognition data, Ruspini data, and letter image recognition data to test the performance of the algorithms they propose. Their results indicate improved and consistent cluster structures compared to a random selection of initial cluster centres.

Sathya and Manavalan [3] used in a study clustering methods such as k-means, improved k-means, fuzzy c-mean (FCM) and improved fuzzy c-mean (IFCM) algorithms. They tested the clustering techniques on different image samples and measured the performance of these algorithms using the segmentation parameters RI, GCE, VoI, and BDE. Their findings indicate that the traditional k-means algorithm completes image segmentation within a shorter period of time but yields inferior results. Moreover, they found that the improved k-means algorithm completes the image segmentation using a smaller number of iterations in comparison to the traditional k-means algorithm. They also compared the conventional FCM and improved FCM algorithms and found that the segmentations performed with improved FCM algorithm yield quality-wise acceptable segmentation using a smaller number of iterations. The improved FCM algorithm also performed better than the other algorithms in terms of performance, accuracy and convergence rate.

In the Augmented K-Means Clustering algorithm proposed in this study, the initial value of the first cluster centre is calculated by dividing the number of bins in the histogram by the number of clusters plus one. This value is also the distance between each cluster centre and the previous one. Thus, the initial values of cluster centres are determined by evenly distributing the cluster centres on the histogram.

Using the proposed algorithm, we segmented a total of 100 (51 benign and 49 malignant) microscopic colon cancer images. We segmented the same images using the Weighted K-Means Clustering algorithm [30] as well and we compared the results. The comparison indicates that the Augmented K-Means Clustering algorithm is better in terms of segmentation success and performance. Furthermore, the Augmented K-Means Clustering algorithm yields more consistent results. Finally, we analyzed the similarity of the images obtained through segmentation, with the original images, and calculated the said similarity by using Histogram Based Similarity algorithm [34]. The results again indicate that the algorithm we developed is better.

This study consists of four sections including the introduction discussing the relevant literature and similar studies, stating the purpose of the study, and offering a short review of the subject matter. In the second section, "Methodology", we explain the experimental design of our study as well as the methods and algorithm we developed. We also provide information about the colon cell images used in the study. In the third section, "Experimental Results", we conduct a performance analysis and articulate its results in several charts and figures. Also in this section, we offer figures comparing

the performance of the algorithm we develop and its counterpart, the Weighted K-Means Clustering algorithm. Finally, we offer a "Conclusions" section to re-emphasize our findings.

2 MATERIAL AND METHODS

2.1 Dataset Selection

In this study, we analyze histologic human colon tissue images that we captured with a camera-mounted microscope. A photograph of the Nikon micrometer microscope slide was also taken during the procedure. The length was calibrated by comparing the photograph of the specimen with the photograph of the Nikon micrometer microscope slide, which was taken under the same magnification. Tissue samples were stained by the Hematoxylin-Eosin technique [35, 36]. Among the 100 images we analyzed, 51 are images of benign colon lesions and 49 belong to malignant colon tumors. These microscopic images were selected at random from the preparations which were stained with Hematoxylin – Eosin, diagnosed and archived at Necmettin Erbakan University, Meram School of Medicine, Department of Pathology.

2.2 K-Means Clustering Algorithm

K-means clustering algorithm [37] is one of the algorithms which was developed by MacQueen in 1967 as a partitioning method. K-Means Clustering algorithm is a clustering algorithm based on measurement of similarity among samples. It performs its function by using several distance measurement methods [38] such as Euclidean, Manhattan and Minkowski. Euclidean distance criterion, one of these methods, is one of most frequently used methods. The K-Means algorithm which was developed four decades ago is today one of the most popular clustering algorithms used in various fields [33]. The development of this algorithm led to the emergence of many similar algorithms in the field. The K-Means algorithm is so popular due to its versatile yet simple and user-friendly problem-solving potential [39]. The stages of the algorithm are as follows:

Suppose, there is a dataset $X=\{X_i, i=1,2,\dots,n\}$ with n elements associated with each channel in the RGB colour space;

Stage 1: The user puts in the cluster number for the dataset and accordingly a cluster $C=\{C_i, i=1,2,\dots,k\}$ gets created, where the cluster centres are assigned random values.

Stage 2: The algorithm calculates the distance between these cluster centre values and the elements of the dataset using the Euclidean distance Eq. (1) and the calculated distance value is assigned to the closest cluster. The same stage is reiterated for all elements of the dataset.

$$d(C_{1,\dots,k}, X_{1,\dots,n}) = \sqrt{\sum_{t=1}^3 (C_{1,\dots,k} - X_{1,\dots,n})^2} \quad (1)$$

Stage 3: The cluster centre (centroid) value $C_{1,\dots,k}=\{Z_i, i=1,2,\dots,m\}$ is calculated by averaging out the values assigned to the elements of the set using the Eq. (2).

$$C_{1,\dots,k} = \frac{1}{m} \sum_{i=1}^m Z_i \quad (2)$$

Stage 4: The Stages 2 and 3 are reiterated until the new cluster values are identical to the previous centroid values.

Here k indicates the number of sets, t is any chrominance channel in RGB space, m is the number of elements at clusters, n is the number of elements at each chrominance channel in RGB space, $C_{1,\dots,k}$ are the centroid values, $X_{1,\dots,n}$ are the elements of the set, $Z_{1,\dots,m}$ are the elements within $C_{1,\dots,k}$ cluster.

2.3 Weighted K-Means Clustering Algorithm

The Weighed K-Means Clustering algorithm [30] works as follows: It uses histogram information concerning the colours in each cluster to calculate the position of the colour cluster centre.

We performed the segmentation processes for our study in RGB colour space. The stages of the algorithm are as follows:

Suppose, there is a dataset $X = \{X_i, i=1,2,\dots,n\}$ with n elements associated with each channel in the RGB colour space;

Stage 1: The user puts in the cluster number for the dataset and accordingly a cluster $C = \{C_i, i=1,2,\dots,k\}$ gets created, where the cluster centres are assigned random values.

Stage 2: The algorithm calculates the distance between these cluster centre values and the elements of the dataset using the Euclidean distance Eq. (1) and calculated distance value is assigned to the closest cluster. The same stage is reiterated for all elements of the dataset.

Stage 3: Unlike the K-Means algorithm, the centroid value $C_{1,\dots,k} = \{Z_i, i=1,2,\dots,m\}$ is calculated by averaging out the elements of a set with the following Eq. (3) which uses the histogram data of the cluster elements.

$$C_k = \frac{1}{\sum_{i=1}^m h_i} \sum_{i=1}^m h_i Z_i \quad (3)$$

Stage 4: The Stages 2 and 3 are reiterated until the new cluster values are identical to the previous centroid values.

Here k indicates the number of sets, t is any chrominance channel in RGB space, m is the number of elements at clusters, n is the number of elements at each chrominance channel in RGB space, h_i is histogram value, $C_{1,\dots,k}$ are the centroid values, $X_{1,\dots,n}$ are the elements of the set, $Z_{1,\dots,m}$ are the elements within $C_{1,\dots,k}$ cluster.

2.4 Augmented K-Means Clustering Algorithm

Augmented K-Means Clustering algorithm determines the number of clusters the same way the k-means algorithms does, by creating a set of clusters, $C = \{C_i, i=1,2,\dots,k\}$. Traditional K-Means Clustering algorithm and Weighted K-Means Clustering algorithm both assign the initial values to cluster centres randomly.

In the algorithm we propose, the initial value assignment to cluster centres is performed by calculating with the following Eq. (4), Eq. (5).

In this algorithm, the stages following the assignment of initial cluster centre values are identical to the Stages 3 and 4 of the Weighted K-Means Clustering algorithm.

$$K_c = \frac{D_h}{k+1} \quad (4)$$

for $i = 1$ to k

$$C_i = K_c(i+1) \quad (5)$$

end for

Here k is the number of clusters, K_c is the cluster center creation coefficient, D_h is the number of bins in histogram, D_h is 256, and C_i is the cluster centre value.

2.5 Histogram Based Similarity Functions

Similarity functions compute a similarity measure between datasets, which are histograms [40]. There are several distance formulas for measuring the similarity of colour histograms. The colour distance formulas measure similarity between images by looking at the perception of colour content [34].

Suppose h and g are two colour histograms. The Euclidean distance between the colour histograms h and g is [34]:

$$h(a,b,c) = N.Prob(A = a, B = b, C = c) \quad (6)$$

$$d(h, g) = \sqrt{\sum_A \sum_B \sum_C (h(a,b,c) - g(a,b,c))^2} \quad (7)$$

Distance Eq. (6), Eq. (7) compares only the identical bins of the respective histograms [34].

Here A, B and C represent the three colour channels (R, G, B) and N is the number of pixels in the image [34].

The calculated similarity value would be 0 if the Histogram-based similarity function is applied to two identical images. In this regard, the greater the similarity between two images is, the closer to 0 the calculated value will be obtained.

3 EXPERIMENTAL RESULTS

In our study, we analyzed 100 microscopic images of colon tissue of which 51 were previously diagnosed to be benign and 49 malignant. First of all, a pathologist marked nuclei in yellow over images of benign and malignant areas; a "gold standard" is created and shown in Fig. 1.

We used 4 separate clustering procedures on the original benign and malignant tissue images. In each procedure we set the number of clusters to one of the following values: $k = \{3, 4, 5, 6\}$ and ran the Weighted K-Means Clustering algorithm and Augmented K-Means Clustering algorithm to effect segmentation. Based on the number of clusters, we used red, green, blue, yellow, dark

green and light blue colours to distinguish each cluster. We provide the clustered images we obtained as a result in Fig. 2, Fig. 3, Fig. 4, and Fig. 5.

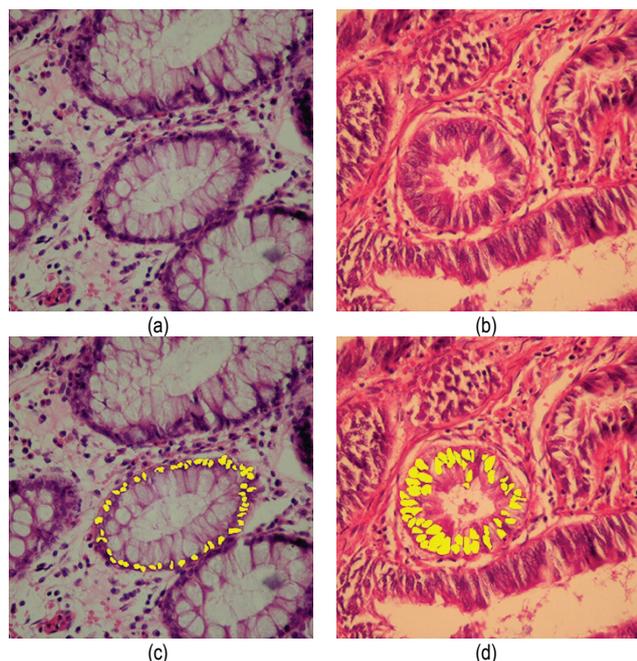


Figure 1 Colon microscopic images: (a) benign microscopic colon image, (b) malignant microscopic colon image, (c) benign gold standard image, (d) malignant gold standard image

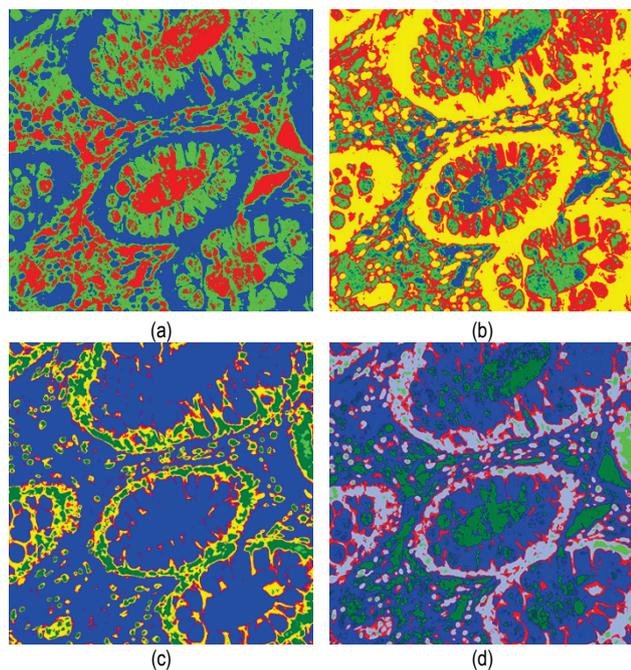


Figure 2 Segmentation by Weighted K-Means Clustering algorithm on benign tissue images: (a) 3 clusters, (b) 4 clusters, (c) 5 clusters, (d) 6 clusters.

Tab. 1 consists of these values for the Weighted K-Means Clustering algorithm and Augmented K-Means Clustering algorithm. For each image, we recorded the number of iterations and processing time. Doing so enabled us to conduct a performance comparison, the results of which can be found in Tab. 2. We also calculated the accuracy of segmentation by comparing the gold standard with the images we obtained as a result of the segmentation. The average time needed for the Weighted K-Means Clustering algorithm and the best

time for the Augmented K-Means Clustering algorithm are 69 sec and 13 sec, respectively.

Table 1 Correlation analysis of number of iterations and process time parameters of segmentation processes repeated three times according to weighted k-means clustering algorithm

Cluster count	1 st and 2 nd process		1 st and 3 rd process		2 nd and 3 rd process	
	Iteration	Time	Iteration	Time	Iteration	Time
3 clusters	0.19	0.20	0.19	0.19	0.18	0.18
4 clusters	0.12	0.12	0.06	0.06	-0.01	-0.01
5 clusters	0.02	0.02	0.19	0.18	-0.02	-0.02
6 clusters	0.01	0.01	-0.12	-0.12	0.03	0.03

As Tab. 1 indicates, during the segmentation process which we repeated three times under identical conditions by reapplying the Weighted K-Means Clustering algorithm, the number of iterations and processing time differed in each reapplication. This is due to the random assignment of the initial cluster values. We detected no correlation between these three applications. Indeed, two distinct applications of the Weighted K-Means Clustering algorithm for segmentation will almost always yield different processing times and iteration outcomes due to the randomization of initial values.

Table 2 Performance analysis of augmented k-means clustering algorithm in comparison to weighted k-means clustering algorithm in terms of number of iterations and process time parameters

Cluster count	Weighted K-Means Clustering Algorithm		Augmented K-Means Clustering Algorithm		Performance (%)	
	Iteration	Time	Iteration	Time	Iteration	Time
3 clusters	43	49	57	51	32.56	4.08
4 clusters	46	49	54	51	17.39	4.08
5 clusters	42	46	58	54	38.10	17.39
6 clusters	31	35	69	65	122.58	85.71

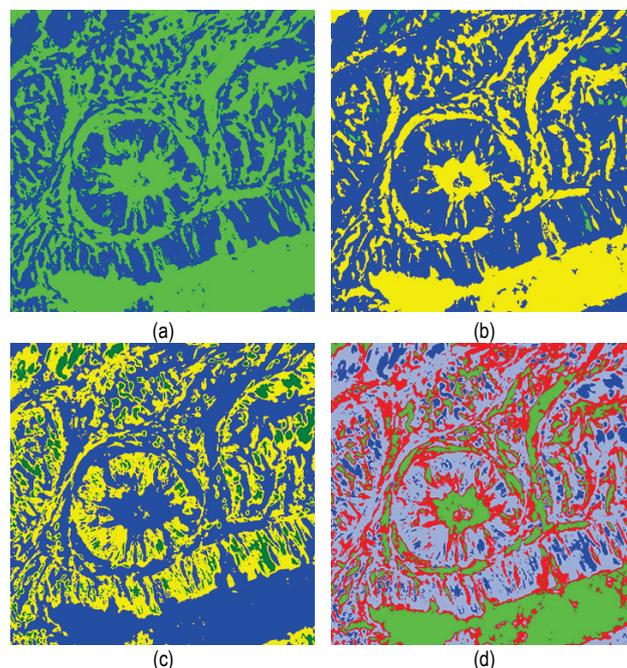


Figure 3 Segmentation by Weighted K-Means Clustering algorithm on malignant tissue images: (a) 3 clusters, (b) 4 clusters, (c) 5 clusters, (d) 6 clusters.

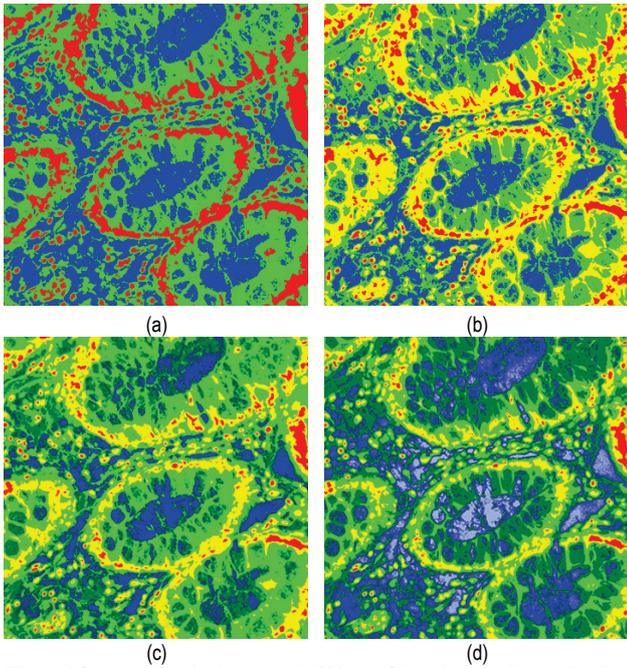


Figure 4 Segmentation by Augmented K-Means Clustering algorithm on **benign** tissue images: (a) 3 clusters, (b) 4 clusters, (c) 5 clusters, (d) 6 clusters.

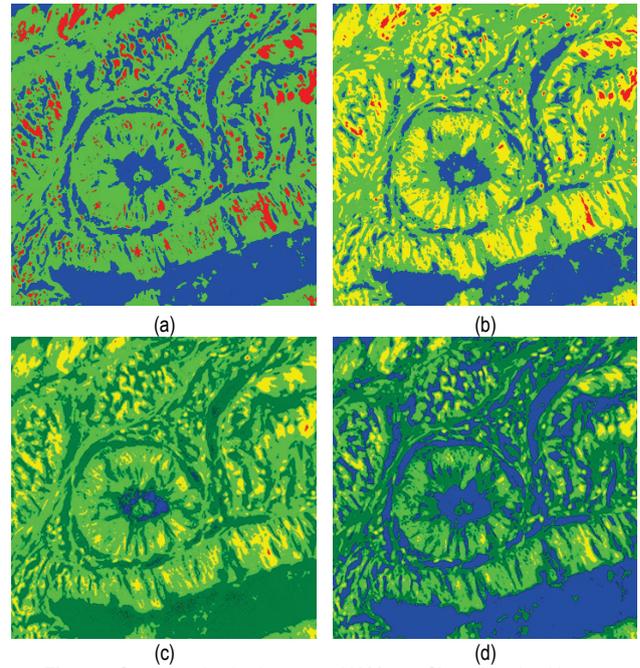


Figure 5 Segmentation by Augmented K-Means Clustering algorithm on **malignant** tissue images: (a) 3 clusters, (b) 4 clusters, (c) 5 clusters, (d) 6 clusters.

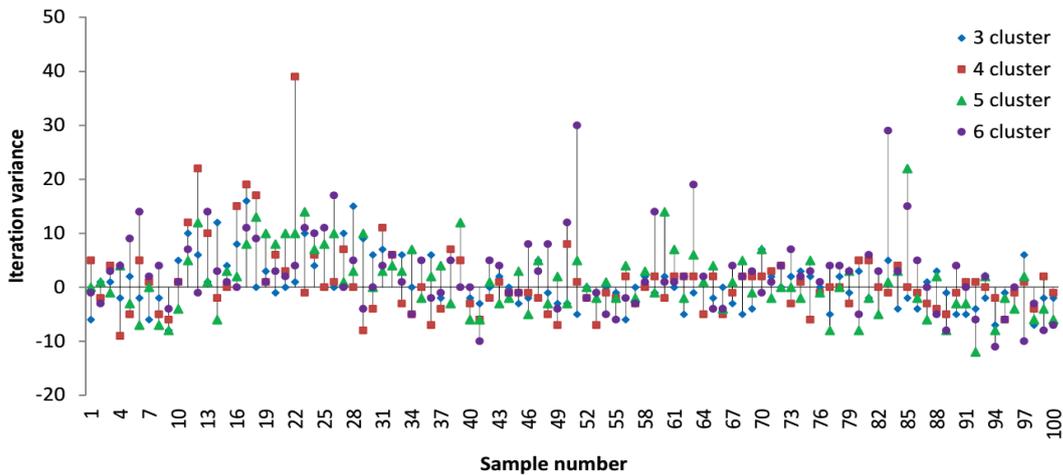


Figure 6 The variance in the number of iterations in segmentation process using Weighted K-Means Clustering and Augmented K-Means Clustering algorithms

We also processed the same images with the Augmented K-Means Clustering algorithm. Here the conditions under which each application took place were also identical and the initial cluster values were assigned deterministically according to the method we developed. As a result, the number of iterations and processing time were identical for each reapplication we performed. Unlike its counterpart, distinct applications of the Augmented K-Means Clustering algorithm for segmentation will always yield the same processing times and iteration outcomes for the same images.

Fig. 6 consists of data concerning the number of iterations variance between the Weighted K-Means Clustering and the Augmented K-Means Clustering algorithms which we used to effect segmentation. We calculated the variance values in question by subtracting the number of iterations the Augmented K-Means Clustering Algorithm needed from the number of iterations needed by the Weighed K-Means Clustering algorithm. On this reading, the positive values in Fig. 6,

for example, are cases in which the number of iterations needed by the Weighted K-Means Clustering algorithm was greater than those needed by its counterpart (and vice versa for the negative values). Fig. 6 is dominated by positive values. In this regard, the Augmented K-Means Clustering algorithm often completes segmentation faster and in less iteration than the Weighted K-Means Clustering algorithm.

Figs. 6 indicate a direct correlation between the variance in the number of iterations.

Tab. 2 contains comparative performance results. It indicates in how many pictures out of 100 an algorithm performed better than its counterpart while using $k = \{3, 4, 5, 6\}$ as the number of clusters. 100 images – both malignant and benign – were segmented using Weighted K-Means Clustering Algorithm and Augmented K-Means Clustering algorithm. The best iteration number in 40 of such images was obtained using Weighted K-Means Clustering algorithm, while Augmented K-Means Clustering algorithm produced such best iteration number

in 59 images (See Tab. 2). The comparison indicates a performance increase while using the Augmented K-Means Clustering algorithm over the Weighted K-Means Clustering algorithm. In order to measure performance and calculate any difference whereof, we used Eq. (8), which takes into account the number of iteration and processing time. This indicated that the Augmented K-Means Clustering algorithm performs better in comparison to the Weighted K-Means Clustering algorithm and as the number of clusters increases, the performance of the Augmented K-Means Clustering algorithm increases as well.

$$P(\%) = \frac{A_2 - A_1}{A_1} 100 \tag{8}$$

Here A_2 is iteration and time value for the Augmented K-Means Clustering algorithm, A_1 is iteration and time value for the Weighted K-Means Clustering algorithm.

Next, we examine the similarity of the segmented images to the originals and investigate how the two algorithms compare. Here we use the Histogram-Based Similarity method, in which similarity values are calculated with the Euclidean convergence model. We offer the results in Tab. 3.

Table 3 Analysis of similarity of segmented images to original image with the method of histogram based similarity

Process	Weighted K-Means Clustering Algorithm				Augmented K-Means Clustering Algorithm			
	3 Cluster	4 Cluster	5 Cluster	6 Cluster	3 Cluster	4 Cluster	5 Cluster	6 Cluster
1 st process	37	34	30	28	63	66	70	72
2 nd process	33	34	37	22	67	66	63	78
3 rd process	31	34	30	34	69	66	70	66
Average	34	34	32	28	66	66	68	72

Each column in Tab. 3 contains the number of segmented images for each reapplication of an algorithm which were more similar to the original than the segmented image produced by the other algorithm. The images segmented by the Augmented K-Means Clustering algorithm are more frequently similar to the original images in comparison to the images segmented by the Weighted K-Means Clustering algorithm.

The histopathologic image of a tissue painted using Hematoxylin and Eosin technique displays three colour groups –purple, pink, and white– and accents thereof. Provided that the segmentation is accurate, the segmented image and the original image will have similar histograms in terms of the distribution of colours, even if the actual images may not look alike. The similarity of histograms between the segmented image and the original image will serve as a criterion to confirm the accuracy of segmentation. That is why histogram-based similarity algorithm was used here. The histogram based similarity algorithm actually serves to verify the accuracy of segmentation, rather than to exhibit the similarity of the images.

Our analyses indicate that the Augmented K-Means Clustering algorithm performs better and more consistently than the Weighted K-Means Clustering algorithm. This is probably due to the random assignment of the cluster centre values by the latter. Furthermore, the images segmented by the Augmented K-Means Clustering algorithm are more frequently similar to the original images and are better segmented in comparison to the gold standard.

4 CONCLUSION

In this study, we used the Augmented K-Means Clustering algorithm and the Weighted K-Means Clustering algorithm to effect segmentation in the RGB colour space of 100 microscopic colon tissue images. We compared the weighted K-Means Clustering algorithm and the Augmented K-Means Clustering algorithm in terms of the number of iterations and processing time. The comparison indicates that the Augmented K-Means

Clustering algorithm performs better in terms of these parameters. To ensure reliability of our findings, we ran each algorithm three times under identical conditions for each image and number of clusters. For the Weighed K-Means Clustering algorithm, which relies on non-linear processing and randomization of the initial cluster values, the segmentation yielded different results in each run. In contrast, each run under identical conditions produced identical results with the Augmented K-Means Clustering algorithm. Using the Histogram-based similarity method, we also calculated the similarity of the images segmented by these two algorithms to the original images. We found that the images segmented by the Augmented K-Means Clustering algorithm are more frequently similar to the originals. Therefore, in microscopic image segmentation applications, the Augmented K-Means Clustering algorithm yields superior results with better segmentation performance. Furthermore, the images segmented by the Augmented K-Means Clustering algorithm are more frequently similar to the originals.

Our findings suggest that the method by which the initial cluster centre values are assigned is very important for the purposes of segmentation performance and consistency. A randomized and non-linear assignment of initial cluster centre values leads to different results in each application of the algorithm. The resulting inconsistency is unacceptable for the purposes of clinically practical diagnostic image analysis. The Augmented K-Means Clustering algorithm we propose not only is more consistent than its counterpart but also performs better and more efficiently, which are important desiderata of clinical applicability.

Acknowledgements

This work was supported by Sakarya University Scientific Research Foundation (Project number: 2010-50-02-026). The authors would like to give their deepest thanks to Sakarya University for financial support. The authors also thank Necmettin Erbakan University, Meram Faculty of Medicine, Department of Pathology for allowing the use of their microscopic colon images.

5 REFERENCES

- [1] Boyle, P. & Levin, B. (2008). *World cancer report 2008*. International Agency for Research on Cancer (IARC), Lyon-France.
- [2] Isik, H., Sezgin, E., & Avunduk, M. C. (2010). A new software program for pathological data analysis. *Comp. in Bio. and Med.*, 40(8), 715-722. <https://doi.org/10.1016/j.compbiomed.2010.06.006>
- [3] Sathya, B. & Manavalan, R. (2011). Image segmentation by clustering methods: Performance analysis. *International Journal of Computer Applications*, 29(11), 27-32. <https://doi.org/10.5120/3688-5127>
- [4] Pham, D. L., Xu, C., & Prince, J. L. (2000). A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2, 315-338. <https://doi.org/10.1146/annurev.bioeng.2.1.315>
- [5] Deng, Y. & Manjunath, B. S. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(8), 800-810. <https://doi.org/10.1109/34.946985>
- [6] Bengtsson, E., Wahlby, C., & Lindblad, J. (2004). Robust cell image segmentation methods. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, 14(2), 157-167.
- [7] Patil, D. D. & Deore, S. G. (2013). Medical image segmentation: A review. *International Journal of Computer Science and Mobile Computing*, 2(1), 22-27.
- [8] Singh, K. K. & Singh, A. (2010). A study of image segmentation algorithms for different types of images. *IJCSI International Journal of Computer Science Issues*, 7(5), 414-417.
- [9] Verma, S., Khare, D., Gupta, R., & Chandel, G. (2013). Analysis of image segmentation algorithms using MATLAB. *Third International Conference on Recent Trends in Information Technology (ICRTIT) / Chennai, India*, 163-172. https://doi.org/10.1007/978-1-4614-3363-7_19
- [10] Adams, R. & Bischof, L. (1994). Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6), 641-647. <https://doi.org/10.1109/34.295913>
- [11] Kancherla, K. & Mukkamala, S. (2013). Early lung cancer detection using nucleus segmentation based features. *Computational Intelligence in Bioinformatics and Computational Biology / Singapore*, 91-95.
- [12] Mat-Isa, N. A., Mashor, M. Y., & Othman, N. H. (2005). Seeded region growing features extraction algorithm; its potential use in improving screening for cervical cancer. *International Journal of the Computer, the Internet and Management*, 13(1), 61-70.
- [13] Pan, Z. & Lu, J. (2007). A Bayes-based region-growing algorithm for medical image segmentation. *Computing in Science Engineering*, 9(4), 32-38. <https://doi.org/10.1109/MCSE.2007.67>
- [14] Pohle, R. & Toennies, K. D. (2001). Segmentation of medical images using adaptive region growing. *Proceedings of SPIE 4322 / San Diego*, 1337-1346.
- [15] Klviinen, H., Hirvonen, P., Xu, L., & Oja, E. (1995). Probabilistic and non-probabilistic hough transforms: overview and comparisons. *Image and Vision Computing*, 13(4), 239-252. [https://doi.org/10.1016/0262-8856\(95\)99713-B](https://doi.org/10.1016/0262-8856(95)99713-B)
- [16] Law, T., Itoh, H., & Seki, H. (1996). Image filtering, edge detection, and edge tracing using fuzzy reasoning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(5), 481-491. <https://doi.org/10.1109/34.494638>
- [17] Sharma, N. & Aggarwal, L. M. (2010). Automated medical image segmentation techniques. *Journal of Medical Physics*, 35(1), 3-14. <https://doi.org/10.4103/0971-6203.58777>
- [18] Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence*, 6, 679-698. <https://doi.org/10.1109/TPAMI.1986.4767851>
- [19] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62-66. <https://doi.org/10.1109/TSMC.1979.4310076>
- [20] Demir, C., Gultekin, S., & Yener, B. (2005). Learning the topological properties of brain tumors. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 2(3), 262-270. <https://doi.org/10.1109/TCBB.2005.42>
- [21] Farjam, R., Soltanian-Zadeh, H., Jafari-Khouzani, K., & Zoroofi, R. A. (2007). An image analysis approach for automatic malignancy determination of prostate pathological images. *Cytometry Part B: Clinical Cytometry*, 72B(4), 227-240. <https://doi.org/10.1002/cyto.b.20162>
- [22] Gunduz-Demir, C. (2007). Mathematical modeling of the malignancy of cancer using graph evolution. *Mathematical Biosciences*, 209(2), 514-527. <https://doi.org/10.1016/j.mbs.2007.03.005>
- [23] Gunduz-Demir, C., Kandemir, M., Tosun, A. B., & Sokmensuer, C. (2010). Automatic segmentation of colon glands using object-graphs. *Medical Image Analysis*, 14(1), 1-12. <https://doi.org/10.1016/j.media.2009.09.001>
- [24] Juang, L. H. & Wu, M. N. (2010). {MRI} brain lesion image detection based on color-converted k-means clustering segmentation. *Measurement*, 43(7), 941-949. <https://doi.org/10.1016/j.measurement.2010.03.013>
- [25] Prasath, P., Perumal, K., Thangavel, K., & Manavalan, R. (2014). A novel approach to gene selection of leukemia dataset using different clustering methods. *Computational Intelligence, Cyber Security and Computational Models, Advances in Intelligent Systems and Computing*, 246, 63-69. https://doi.org/10.1007/978-81-322-1680-3_7
- [26] Yogesan, K., Jrgensen, T., Albrechtsen, F., Tveter, K. J., & Danielsen, H. E. (1996). Entropy-based texture analysis of chromatin structure in advanced prostate cancer. *Cytometry*, 24(3), 268-276. [https://doi.org/10.1002/\(SICI\)1097-0320\(19960701\)24:3<268::AID-CYTO10>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0320(19960701)24:3<268::AID-CYTO10>3.0.CO;2-O)
- [27] Boykov, Y. & Funka-Lea, G. (2006). Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision*, 70(2), 109-131. <https://doi.org/10.1007/s11263-006-7934-5>
- [28] Felzenszwalb, P. F. & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2), 167-181. <https://doi.org/10.1023/B:VISI.0000022288.19776.77>
- [29] Shi, J. & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8), 888-905. <https://doi.org/10.1109/34.868688>
- [30] Albayrak, S. (2001). Color quantization by modified k-means algorithm. *Pakistan Journal of Applied Science*, 1(4), 508-511.
- [31] Albayrak, S. & Karslıgil, M. Y. (2001). The color clustering in color images with weighted k-means method. *9th Signal Processing and Application Congress / KKTC 2001*, 609-614.
- [32] Mignotte, M. (2008). Segmentation by fusion of histogram-based-means clusters in different color spaces. *Trans. Img. Proc.*, 17(5), 780-787. <https://doi.org/10.1109/TIP.2008.920761>
- [33] Khan, S. S. & Ahmad, A. (2004). Cluster center initialization algorithm for k-means clustering. *Pattern Recognition Letters*, 25(11), 1293-1302. <https://doi.org/10.1016/j.patrec.2004.04.007>
- [34] Jeong, S. (2001). Histogram-based color image retrieval. Stanford University, <http://scien.stanford.edu/pages/labsite/2002/psych221/projects/02/sojeong/>
- [35] Ekicioglu, G., Ozkan, N., & Salvaazar, E. (2005). Hematoksilen-eozin (hematoxylin-eosin). *Aegean Pathology Journal*, 2(2), 58-61.

- [36] Lillie, R. & Fullmer, H. (1976). *Histopathologic technic and practical histochemistry*. McGraw-Hill Inc. 4th ed., US.
- [37] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematical Statistics and Probability / California*, 281-297.
- [38] Li, X., Yang, J., Wang, Q., Fan, J., & Liu, P. (2008). Research and application of improved k-means algorithm based on fuzzy feature selection. *Fifth International Conference on Fuzzy Systems and Knowledge Discovery / Shandong*, 401-405. <https://doi.org/10.1109/FSKD.2008.451>
- [39] Kumar, P. & Wasan, S. K. (2010). Comparative analysis of k-means based algorithms. *International Journal of Computer Science and Network Security*, 10(4), 281-297.
- [40] Zweng, A., Rittler, T., & Kampel, M. (2011). Evaluation of histogram-based similarity functions for different color spaces. *14th International Conference on Computer Analysis of Images and Patterns / Spain*, 455-462. https://doi.org/10.1007/978-3-642-23678-5_54

Contact information:

Ulaş YURTSEVER, PhD student
(Corresponding author)
Sakarya University, Institute of Natural Sciences,
Computer and Information Engineering,
54187, Sakarya, Turkey
E-mail: ulas@sakarya.edu.tr

Hayrettin EVIRGEN, PhD, Assistant professor
İstanbul University,
Faculty of Open and Distance Education,
İstanbul, Turkey
E-mail: hayrettin.evirgen@auzefmail.istanbul.edu.tr

Mustafa Cihat AVUNDUK, PhD, Professor
Necmettin Erbakan University,
Faculty of Meram Medical,
Department of Pathology,
Konya, Turkey
E-mail: mcavunduk@hotmail.com