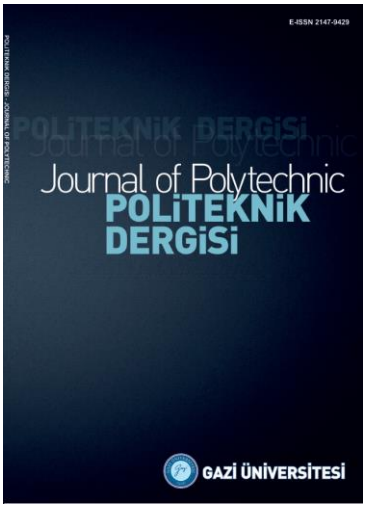

POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.gov.tr/politeknik

A novel probabilistic nuclei segmentation algorithm for H&E stained histopathological tissue images

H&E ile boyanmış histopatolojik doku imgeleri için yeni bir olasılıksal hücre çekirdeği bölütleme algoritması

***Yazar(lar) (Author(s))**: Faruk SERIN[1] , Metin ERTURKLER[2], Mehmet GUL[3]*

ORCID[1]: 0000-0002-1458-4508

ORCID[2]: 0000-0003-0195-4028

ORCID[3]: 0000-0002-1374-0783

To cite to this article (Bu makaleye şu şekilde atıfta bulunabilirsiniz): Serin F., Erturkler M. ve Gul M., "A novel probabilistic nuclei segmentation algorithm for H&E stained histopathological tissue images", ***Politeknik Dergisi***, 23(1): 7-17, (2020).

Erişim linki (To link to this article): http://dergipark.gov.tr/politeknik/archive

DOI: **10.2339/politeknik.464541**

A Novel Probabilistic Nuclei Segmentation Algorithm for H&E Stained Histopathological Tissue Images

Araştırma Makalesi / Research Article

Faruk SERIN[1*], Metin ERTURKLER[2], Mehmet GUL[3]

[1]Department of Computer Engineering, Faculty of Engineering, Munzur University, Tunceli, Turkey

[2]Department of Computer Engineering, Faculty of Engineering, Inonu University, Malatya, Turkey

[3]Department of Embryology and Histology, Faculty of Medicine, Inonu University, Malatya, Turkey

(Received/Geliş: 27.09.2018 ; Accepted/Kabul : 28.02.2019)

ABSTRACT

In this study, we propose a novel, fast and accurate segmentation algorithm to segment nuclei in H&E stained histopathological tissue images. The proposed algorithm does not require pre-processing, post-processing, and any manual parameter or threshold. The algorithm utilizes probabilistic and statistical properties of the pixels' color value in the images with RGB color space, and determines whether pixels are a part of any nuclei or not by using an automatically calculated threshold value. The algorithm provides time efficiency and reduced overall cost in the segmentation. Two more algorithms are also proposed to distinguish nuclei cluster from the other clusters obtained by K-means, and eliminate false positives in nuclei cluster, which are not nuclei. In order to compare and evaluate the performance of the proposed segmentation algorithm in terms of time and cost efficiency, K-Means is preferred because of its common usage. Expert evaluation is declared as ground truth for determining the accuracy of the results. The experiments are performed on 60 healthy and 60 damaged kidney, and 60 healthy and 60 damaged liver tissue images. The evaluations show that the proposed algorithm can effectively segment nuclei. The comparison results also demonstrate that the deviation between proposed algorithm and the expert is 2%, while the deviation between K-Means and expert is 5%.

Keywords: Image segmentation, medical image processing, clustering methods, pattern recognition.

H&E ile Boyanmış Histopatolojik Doku İmgeleri için Yeni Bir Olasılıksal Hücre Çekirdeği Bölütleme Algoritması

ÖZ

Bu çalışmada, H&E boyalı histopatolojik doku imgelerindeki hücre çekirdeklerini bölütlemek için yeni, hızlı ve doğru bir bölütleme algoritması önerilmiştir. Önerilen algoritma ön işlem, son işlem, herhangi bir manuel parametre veya eşik değeri gerektirmez. Algoritma, RGB renk uzayında olan imgelerdeki piksellerin renk değerinin olasılıksal ve istatistiksel özelliklerini kullanır ve piksellerin herhangi bir çekirdeğin bir parçası olup olmadığını otomatik olarak hesaplanan eşik değeri kullanarak belirler. Algoritma, zaman verimliliği sağlar ve bölütleme genel maliyetini düşürür. Ayrıca, K-ortalama sonucu elde edilen kümeler içerisinden hücre çekirdeklerini içeren kümenin belirlenmesi ve hücre çekirdekleri kümesi içerisinde bulunan ancak hücre çekirdeği olmayan yanlış pozitiflerin elimine edilmesi için iki algoritma daha önerilmiştir. Önerilen bölütleme algoritmasının zaman ve maliyet verimliliği açısından performansını karşılaştırmak ve değerlendirmek için, yaygın kullanımı nedeniyle K-ortalama bölütleme algoritması tercih edilmiştir. Sonuçların doğruluğunu belirlenmesi için uzman değerlendirmesi baz alınmıştır. Deneyler 60 sağlıklı ve 60 hasarlı böbrek ile 60 sağlıklı ve 60 hasarlı karaciğer görüntüsü üzerinde gerçekleştirilmiştir. Değerlendirmeler, önerilen algoritmanın çekirdekleri etkili bir şekilde bölütleyebildiğini göstermektedir. Karşılaştırma sonuçları ayrıca önerilen algoritma ile uzman arasındaki sapmanın %2 olduğunu, K-Ortalama ve uzman arasındaki sapmanın ise %5 olduğunu göstermektedir.

Anahtar Kelimeler: Görüntü bölütleme, medikal görüntü işleme, kümeleme yöntemleri, örüntü tanıma.

1. INTRODUCTION

In histopathological analysis, the tissue samples are prepared by performing routine technical procedures which are fixation, dehydration, clearing, infiltration, embedding, sectioning, and staining respectively [1,2]. Staining procedure aims to reveal different tissue structures by staining it with different colors. Hematoxylin-Eosin (H&E) is the most common staining technique and many pathologists believe that H&E will continue to be common practice over the next 50 years [3,4]. Hematoxylin stains nuclei with blue color specification, while Eosin stains other structures such as cytoplasm, connective tissue, vascular lumen, etc. with white and pink color specification. Histopathological images can be obtained by different imaging techniques [2,5] depending on usage purposes. Fast slide scanners are usually used to generate digital histopathological

**Sorumlu Yazar (Corresponding Author)*
e-posta : bmfarukserin@gmail.com

images that contain relevant information about the specimen at a microscopic imaging.

An expert evaluates digital histopathological images manually or by a Computer Assisted Diagnosis (CAD) system. In practice, experts visually examine the tissue sections through a microscope. However, this approach is slow, time consuming [6] and error-prone due to inexperience and subjectivity of experts. It is also presented in [7–9] that the experts make joint decision between 61% and 73%.

CAD systems have been used in histopathology besides other medical fields to provide quantitative data to experts in diagnosis process [3,10–23]. Furthermore, these data acquired by a CAD system may provide more confidence in decision-making process during a diagnosis. CAD systems also reduce the workload, and speed up the diagnosis and treatment process. Thanks to recent advances in CAD systems, different various disease detection and grading application have been proposed, including counting of ovary follicles [24], compute areas and volume of the scar cardiac tissue [19], diagnosing melanocytic and non-melanocytic skin lesions [25], locating and qualification of fatty and necrosis area of liver tissues [21–23], detection and analysis of cancer such as breast [26], prostate [27], lung [28].

Segmentation is an important initial step of many CAD systems, and the success of the CAD systems largely depends on the quality of the segmentation algorithm. A segmentation algorithm aims to separate interested tissue structure from others. However, it is not an easy task due to the complex nature of histopathological images and the variability in the sample preparation, staining and image acquisition process. In many histopathological image analysis systems, segmentation mainly focuses on identification of nuclei owing to its major and distinct properties. The nuclei can be identified by using general or specific image segmentation algorithms.

Well-known image segmentation algorithms can be categorized based on threshold, edge, region, and cluster in the literature. Threshold based segmentation assumes that certain structures have significantly different intensities than the background or other structures. It applies a fixed threshold value to image globally or locally. The threshold value can be depending on the global or local features of the image such as color intensities, histogram, and statistical properties of the features. The threshold value can be determined manually or calculated by a method such as Otsu [29]. Edge based segmentation algorithms are performed on especially gray level images by detecting strength intensity changes in pixels. The first or second order derivative operators such as Prewitt, Sobel, Canny, Test, Zero–Crossings and Laplacian are used to identify the pixels changing. These pixels constitute of the boundary of components. Region based segmentation algorithms aim to divide the image into homogeneous sub-regions considering similarity criteria such as a thresholded color value, equality of gray level, or relationship between any features based on histogram, color, etc. Region Growing [30], Splitting and Merging [31] are well known algorithms of this category. Clustering based segmentation algorithms such as K-Means, Fuzzy C-Means and Unsupervised Fuzzy C-Means separate images into different homogeneous sub-images called cluster. K-means [13,22,23,32] and Fuzzy C-Means [33,34] divide image into n clusters. The determination of number and initial pixels of clusters plays essential role in success of the segmentation.

Specific nuclei segmentation algorithms have also been presented in the literature. H. Kong et al. proposed to classify pixels of histopathological images into cell and extra-cellular clusters by using color-texture properties instead of color intensities. The color-texture at each pixel is extracted by using local Fourier transform from the most discriminant color space that is optimized to be a linear combination of the original RGB color space [35]. X. Zhang et al. developed a segmentation method to describe cell by using Gaussian-based hierarchical voting and repulsive balloon model [36]. Y. Xu et al. proposed the multiple clustered instance learning to classify, segment and cluster medical images [37]. S. Wienert et al. suggested contour based cell detection and segmentation method utilizing minimal priori information and avoiding segmentation bias related to shape features. [38]. Y. Al-Kofahi et al. proposed a novel segmentation method consisting of various ideas. In first step of the method, foreground is separated from histopathological images by using graph-cut based binarization. After that, seed point of nuclei is obtained by multi-scale Laplacian of Gaussian filtering that is restrained by distance-map-based adaptive scale selection. These points are then used to perform initial segmentation improved by using second graph-cut based algorithm including the alpha expansions and graph coloring methods [39].

The well-known image segmentation and specific histopathological image segmentation algorithms require pre-processing, post-processing, algorithm specific parameter or threshold value determined manually. These increase the mathematical calculations naturally. However, the nuclei segmentation can be performed without these operations to achieve relatively reduced overall operational cost, if the concern is to segment cell nuclei in H&E stained histopathologic images, which is major structure of a tissue especially in cancer researches.

Two main contributions are presented in this study. The first contribution is proposed Probabilistic Nuclei Segmentation Algorithm. The algorithm is novel, fast and accurate. In addition, it does not require pre-processing, post-processing, and any parameter or threshold value determined manually, to identify nuclei in H&E stained histopathological images. The algorithm utilizes probabilistic and statistical properties of the pixels' color value in the images with RGB color space, and determines whether pixels are a part of any nuclei or

not by using an automatically calculated threshold value. The probabilistic and statistical operations are focused on blue channel of the RGB color space due to the fact that the cell nuclei are stained with blue in the H&E stained histopathological images. This approach provides time efficiency and reduced overall cost in nuclei segmentation process. Two more algorithms are proposed as second contribution to distinguish cluster containing nuclei from the other clusters obtained by K-means, and eliminate false positives in nuclei cluster, which are not nuclei.

We preferred to compare our proposed algorithm with a well-known segmentation algorithm, which has been used in histopathological image segmentation, to evaluate the performance of the proposed nuclei segmentation algorithm in terms of time and cost efficiency as well as accuracy of results. For this, K-means is considered because of its common usage [12,13,22,23,32,40,41]. The accuracy of an algorithm for nuclei segmentation is to locate and identify all nuclei correctly in histopathological image. Expert evaluation is declared as ground truth for determining the accuracy of the proposed algorithm.

The experiments performed on 60 healthy and 60 damaged kidney, and 60 healthy and 60 damaged liver tissue images reveal that the proposed algorithm is effective to obtain accurate nuclei segmentation and cell counting. In addition, the experiments show that the results of the proposed algorithm are closer to expert's evaluation than the results of K-Means. It is also shown that the proposed algorithm is faster than K-Means.

The rest of this paper is organized as follows. Section 2 highlights the requirements of the histopathological image segmentation algorithms and introduces K-Means. Nuclei distinguishing, and false positive removal algorithm are also explained in Section 2. The proposed segmentation algorithm is presented and analyzed in section 3. Section 4 describes dataset, evaluations and discussions of experiments. Conclusions are summarized in section 5.

2. CELL NUCLEI SEGMENTATION

The well-known histopathological image segmentation algorithms generally require pre-processing, post-processing, algorithm specific parameters or threshold value determined manually as stated in introduction. Watershed needs to convert image color space from RGB to gray-scale as a pre-processing step. Similarly, thresholding requires transforming color space from RGB to binary or gray-scale in pre-processing step. K-Means, Watershed, Region Growing and Fuzzy C-Means algorithms segment image into different clusters, but the algorithms does not identify the cluster containing nuclei. Thus, the algorithms include a post-processing step. In addition, K-Means needs parameters to determine the number of cluster and initial location of each cluster. Region Growing requires also number of seed and initial position of each seed. In threshold-based segmentation algorithms, determination of a threshold value parameter is necessary as well.

The proposed algorithm is compared to K-Means Segmentation Algorithm since it is widely used in histopathological image segmentation [12,13,22,23, 32,40,41].

2.1. K-Means Segmentation Algorithm

K-Means Segmentation Algorithm divides image into homogeneous sub-images as clusters. Thus, it is essential to determine number of clusters for clear segmentation. The number of clusters can be determined manually or computationally [42]. If number of clusters is predictable, it is determined manually to avoid raising the computational cost. K-Means Segmentation Algorithm is expressed step by step as follows:

1. Number of clusters is determined.
2. Initial centroid of clusters is selected.
3. Distances between pixels and centroids are calculated by using Euclidian distance as in (1).

$$d_{i,j} = ||p_i - \mu_j|| \quad (1)$$

where p_i is location of i^{th} pixel, i = [1, 2, 3, …, the number of pixels], μ_j is centroid of j^{th} cluster, j = [1, 2, 3, …, the number of clusters k], $d_{i,j}$ is Euclidian distance between location of i^{th} pixel and centroid of j^{th} cluster.

4. Each pixel is assigned to nearest cluster.
5. Centroid of clusters is recalculated as in (2).

$$\mu_j = \frac{1}{n_j} \sum_{l=1}^{n_j} p_l^j \quad (2)$$

where n_j is number of pixels in j^{th} cluster, p_l^j is location of l^{th} pixels in j^{th} cluster.

6. Steps 3, 4 and 5 are repeated as long as the centroid of clusters shifts. This repetition can be terminated when the amount of shift is smaller than a predefined threshold value.

Cost function of K-Means J is calculated by summing of squares of Euclidian distances between each pixel and centroid of cluster that contains the pixel as in (3).

$$J = \sum_{j=1}^{k} \sum_{l=1}^{n_j} \left\| p_l^j - \mu_j \right\|^2 \quad (3)$$

A histopathological liver tissue image shown in Figure 1(a) is used to demonstrate the segmentation results of K-Means with 5 clusters. The clusters results of segmentation are shown in Figure 1(b-f). The structures in histopathological image are assigned to different clusters as seen in Figure 1. In this paper, we propose an algorithm to distinguish nuclei cluster from these different clusters automatically, since K-Means cannot distinguish nuclei cluster itself without post-processing.

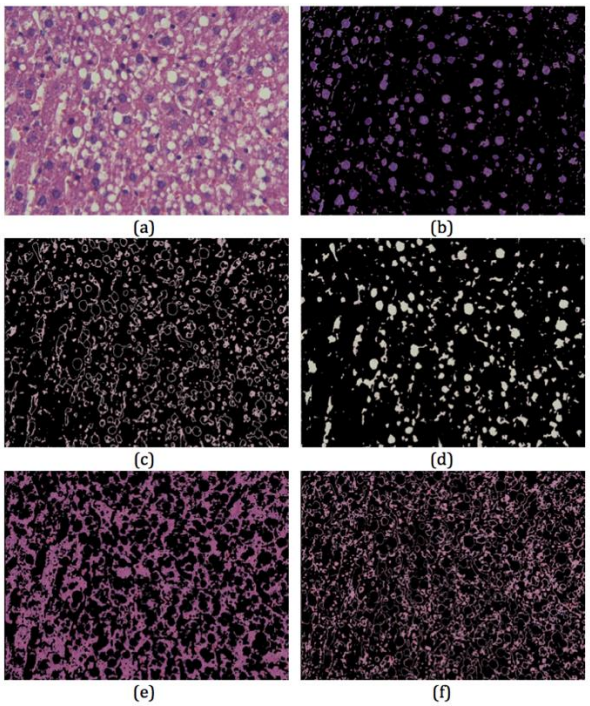

Figure 1. A histopathological liver tissue image and its K-Means segmentation result with 5 clusters (a) Liver image. (b) Nuclei. (d) Fat vacuoles, vascular lumens and connective tissue. (c), (e), and (f) Cytoplasm.

2.2. Distinguishing Nuclei Cluster

The segmented structures are not identified after cluster-based segmentation algorithms such as K-Means. Thus, it is required to distinguish the cluster containing the nuclei from the other clusters. We propose to take advantage of average of structure's color intensity value for identification. First, clusters are converted to gray level images and average of structure's gray level intensity value is calculated by considering only pixels of white foreground structures without pixels of black background area. Second, cluster with minimum average is identified as the nuclei cluster because structures stained with blue color have lower intensity value than the other image structures in H&E stained histopathological tissue images. Distinguishing of the nuclei cluster is formulated in (4), (5) and (6).

$$S_j = \sum_{l=1}^{n_j} I_j\left(p_l^j\right) \tag{4}$$

$$C_j = \frac{S_j}{n_j} \tag{5}$$

$$N = minimum(C) \tag{6}$$

where n_j is number of pixels in j^{th} cluster, p_l^j is location of l^{th} pixels in j^{th} cluster, I_j is gray level image of j^{th} cluster, S_j is total gray level intensity of I_j, C_j is average gray level intensity of I_j, N is nuclei cluster binary image.

The obtained nuclei cluster after applying the proposed distinguishing algorithm to clusters in Figure 1(b-f) is shown in Figure 2(a). As seen in Figure 1(b) and Figure 2(a), nuclei cluster contains false positive tiny points that are not nuclei. These points have to be removed from cluster to segment nuclei accurately.

2.3. False Positive Removal

Tiny points in blue color specification can be seen in the H&E stained histopathological images when the images are examined in detail. These misleading points can occur due to the complex nature of histopathological tissue and the variability in the sample preparation, staining and image acquisition process. Thus, these points cause to appear false positive points (FPP) in nuclei cluster. These FPP are in fact not part of any nuclei. The FPP is required to be removed from nuclei cluster for preventing identification of misleading points as cell nuclei. A false positive removal algorithm is proposed to remove the FPP.

Area of FPP is considerably smaller than a nucleus area. Thus, area thresholding is applied to nuclei cluster to remove FPP in binary nuclei cluster. The threshold can be determined manually depending on nature and resolution of image or automatically. In automatic thresholding, first *r* the ratio between number of pixels in foreground of binary nuclei cluster and resolution of image (total pixels count of the image) is computed as in (7). The calculated ratio *r* is used as circle radius. Then, the area of circle of radius *r* is calculated as in (8) and this area is used as area threshold Ꞇ. Finally, the component whose area is smaller than Ꞇ is removed from binary nuclei image as in (9).

$$r = \frac{\psi}{\varphi} \tag{7}$$

$$Ꞇ = \pi r^2 \tag{8}$$

$$Ƒ_i = \begin{cases} 1, & A_i \geq Ꞇ \\ 0, & A_i < Ꞇ \end{cases} \tag{9}$$

where ψ is the resolution of image, which is number of pixel in image; φ is the number of pixels in foreground of binary nuclei image; *r* is the ratio between the number of pixels in foreground of binary nuclei image and resolution of image; Ꞇ is the area of circle of radius *r*; A_i is the area of i^{th} foreground component, which is the number of pixels in the component; $Ƒ_i$ is false positive removal function of i^{th} component. The component is removed from image if $Ƒ_i$ equals to 0. The solution for overlapping nuclei is proposed in [11].

The proposed false positive removal algorithm is applied to nuclei cluster in Figure 2(a). The obtained nuclei are shown in Figure 2(b) where FPP are removed.

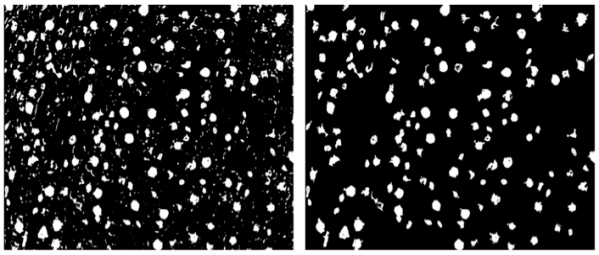

Figure 2. Segmentation result of Figure 1(a) by K-Means Segmentation Algorithm (a) Binary nuclei cluster that is result of distinguishing algorithm. (b) Nuclei that is result of false positive removal algorithm.

3. THE PROPOSED SEGMENTATION ALGORITHM

Tissue staining technique plays an important role in selection and success of segmentation algorithms. Many pathologists believe H&E staining will maintain its dominance in practice over the next 50 years [3,4]. When histopathological tissue sections are stained with H&E, cell nuclei are typically stained with blue color and its specifications in the images with RGB (Red, Green, and Blue) color space [5]. Thus, blue color value of nuclei is generally greater than their red and green color value. The proposed algorithm aims to separate nuclei from the other structures in H&E stained histopathological tissue images on the strength of these truths. The overall schematic of the proposed algorithm is shown in Figure 3.

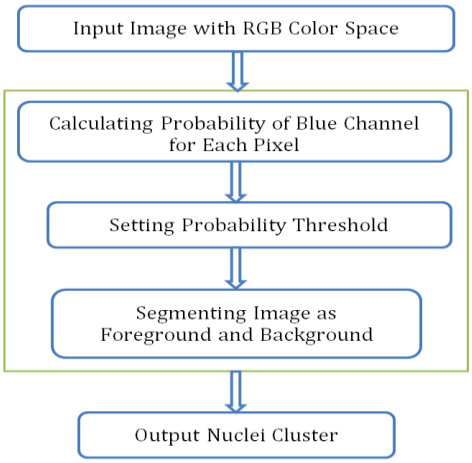

Figure 3. The overall schematic of the proposed algorithm

The proposed algorithm processes an input image with RGB color space depending on color specification the image contains. The regions that contain blue color specifications are determined as foreground (nuclei) while the rest of the image is determined as background after the proposed algorithm processes each pixel in the image. As shown in Figure 4, the output of foreground is white, and the output of background is black.

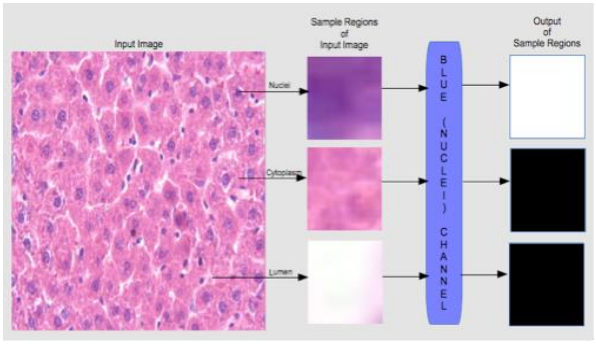

Figure 4. The sample input and output regions for the proposed algorithm working on blue channel

The proposed algorithm firstly calculates the ratio of blue (B) value for each pixel as in (10). This ratio represents the probability of being a part of any nuclei for the pixels. By utilizing these probabilities, the threshold value t is set as in (11), (12) and (13) to decide if the pixel is a part of any nuclei or background. The decision function is presented in (14).

$$p_i = \frac{b_i}{(r_i + g_i + b_i)} \tag{10}$$

where r_i, g_i, and b_i are red, green, and blue values of i^{th} pixel in RGB color space respectively; p_i is the ratio of the blue value that represents the probability of being a part of any nuclei for i^{th} pixel.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{11}$$

$$\sigma = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(p_i - \mu)^2} \tag{12}$$

$$t = \mu + \sigma \tag{13}$$

where n is number of pixels, μ is the mean of the probabilities, σ is standard deviation of probabilities.

$$f_i = \begin{cases} 1, & p_i \geq t \\ 0, & otherwise \end{cases} \tag{14}$$

The pixel is a part of any nuclei, if decision function f_i equals to 1 for i^{th} pixel; otherwise the pixel belongs to the background. If set of images is homogeneous and threshold value is predictable, t can be calculated once or defined manually. This reduces workload, and speeds up process. However, various processes and factors such as faults in any histopathological procedure and imaging generally make the dataset heterogeneous. Thus, t is calculated as in (11-13). The homogeneity means that the similar structures in image dataset are in the similar color. The pseudocode of the proposed algorithm is given in Algorithm 1.

Algorithm 1. Probabilistic Nuclei Segmentation Algorithm

Input: RGB image $I(x,y,z) = \{u \mid 255 \geq u \geq 0 \wedge u,x,y \in \mathbb{N} \wedge z = \{0,1,2\}\}$

Output: Binary image $B(x,y) = \{0,1\}$ and $x,y \in \mathbb{N}$

Variables:
v is vertical size of I
h is horizontal size of I
I(i,j,0) is red (R) color intensity of I
I(i,j,1) is green (G) color intensity of I
I(i,j,2) is blue (B) color intensity of I
p is probability
μ is mean
σ is standart deviation
t is threshold probability
prob is temporary variable
n is temporary pixel counter variable
δ is temporary variable

Algorithm:

1 prob=0;
2 n=0;
3 δ=0;
4 **for** i = 0; i < v; i=i+1 **do**
5 **for** j = 0; j < h; j=j+1 **do**
6 $p_{i,j} = I(i,j,2) / (I(i,j,0) + I(i,j,1) + I(i,j,2))$;
7 $prob = prob + p_{i,j}$;
8 **end**
9 **end**
10 $n = h \times v$;
11 $\mu = prob / n$;
12 **for** i = 0; i < v; i = i + 1 **do**
13 **for** j = 0; j < h; j = j + 1 **do**
14 $\delta = \delta + (p_{i,j} - \mu)^2$;
15 **end**
16 **end**
17 $\sigma = \sqrt{\delta/(n-1)}$;
18 $t = \mu + \sigma$;
19 **for** i = 0; i < v; i=i+1 **do**
20 **for** j = 0; j < h; j = j + 1 **do**
21 **if** $p_{i,j} \geq t$ **then**
22 $B(i,j) = 1$;
23 **end**
24 **else**
25 $B(i,j) = 0$;
26 **end**
27 **end**
28 **end**

3.1. Time Complexity of The Probabilistic Nuclei Segmentation Algorithm

Table 1. Computation time of the proposed algorithm

Step	Computation	Time (CC)
1	T	1
2	T	1
3	T	1
4	vA+(v+1)C+ (v+1)T	3v+2
5	v(hA+(h+1)C+ (h+1)T)	3vh+2v
6	vh(T+D+2A)	4vh
7	vh(T+A)	2vh
8	-	-
9	-	-
10	T+M	2
11	T+D	2
12	vA+(v+1)C+ (v+1)T	3v+2
13	v(hA+(h+1)C+ (h+1)T)	3vh+2v
14	vh(T+A+2S+M)	5vh
15	-	-
16	-	-
17	T+R+D+S	4
18	T+A	2
19	vA+(v+1)C+ (v+1)T	3v+2
20	v(hA+(h+1)C+ (h+1)T)	3vh+2v
21	vhC	vh
22	vhT	vh
23	-	-
24	-	-
25	vhT	vh
26	-	-
27	-	-
28	-	-
Total		**22vh+15v+19**

Time complexity is important to evaluate the algorithm efficiency. Thus, time complexity of proposed Probabilistic Nuclei Segmentation Algorithm given in Algorithm 1 is calculated step by step in Table 1. The differences between the computation times of mathematical operations such as addition (A), subtraction (S), multiplication (M), division (D), comparison (C), square root (R) and assignment (T) are ignored. The computation time of the mathematical operations is denoted as A, S, M, D, C, R and T respectively. Computation time of each operation is accepted as 1 clock cycle (CC). For example, a=b/f+d+e is denoted as T+D+2A, which means there are 1 assignment, 1 division and 2 addition. The total computation time of the expression is equal to 4 CC.

Step 22 and 25 do not operate together. Only one of them is operated and total computation time of these two steps

is *vh*. The worst case, best case and average case of the proposed algorithm is the same as in (15).

$$T = 22vh + 15v + 19 = O(vh) = O(n) \quad (15)$$

As seen in (15), the time complexity of the proposed algorithm is O(n) where n is number of pixels in the image.

4. EXPERIMENTS

4.1. Datasets

In this study, histopathological images of liver and kidney tissues were obtained by the expert at Inonu University and these images were used for experiment. Liver tissue samples were removed from 150-180g male Wistar Albino Rats grouped as healthy (control) damaged. The damaged group were injected with olive oil and 1 ml/kg/day CCl_4 a xenobiotic used in the generation of experimental liver damage, and causes injury on liver by increasing generation of free radicals, decreasing the activities of antioxidant enzymes, and inducing lipid peroxidation. Kidney tissue samples were removed from left kidney of 250-280g male Sprague Dawley Rats grouped as damaged and healthy. 100 mg/kg/day the intraperitoneal Aluminium Chloride ($AlCl_3$) was injected to damaged group. The rats had been held in 22-24 ºC rooms, the seasonal daylight period, and fed with standard rat chow pellets and tap water. The rats were slept using Ketamine (Ketalar, Parke-Davis, Eczacıbaşı, Turkey) and Xylazine HCl (Alfazyne %2; Alfasan, Woerden, Netherlands) anesthesia and tissue samples were taken. The samples determined with 10% formaldehyde were embedded into paraffin block through routine histologic follow-up procedures. The sections were cut with a microtome and stained with H&E. The images were taken from these sections by Leica DFC280 light microscopy and Leica Q Win (Leica Microsystems Imaging Solutions, Cambridge, UK) image analysis system.

Kidney images have 1920x2560 and liver images have 768x1024 resolution. Eight of the obtained 240 histopathological images are shown in Figure 5 where there are three components with blue, pink and white color. Blue ones represent nuclei whereas pink and white ones represent other tissue structures. Dataset consists of 60 healthy and 60 damaged kidney, and 60 healthy and 60 damaged liver tissue images.

4.2. Evaluation

Segmentation of cell nuclei is a crucial step in automatic analysis of histopathological images. In the experiment, K-Means [43] and the proposed algorithm are used for segmentation. In preprocessing step for K-Means, RGB images are converted to L*a*b* color space image to reduce color space from three channels (R-G-B) to two channels (a*-b*). 'L*' layer of the L*a*b* space indicates a luminosity, 'a*' layer indicates chromaticity (red-green), and 'b*' layer indicates chromaticity (blue-yellow). K-Means is applied to 'a*' and 'b*' layers in which all of the color information is. Number of clusters for K-Means is selected as 5. Proposed distinguishing algorithm is applied to distinguish nuclei cluster from other clusters. False positive removal algorithm is used to remove non-nuclei components from nuclei cluster.

The segmentation results of K-Means and the proposed algorithm for eight liver and kidney images are shown in Figure 5. The nuclei segmented by the proposed algorithm and K-Means can be counted by any connected component labeling method such as [44]. The expert also counted the nuclei in images manually. The number of nuclei segmented by K-Means, the proposed algorithm, and expert is illustrated in Table 2 for quantitative comparison. As shown in Table 2, the number of the cell counted by the proposed algorithm is closer to expert evaluation than K-means.

The number of nuclei segmented by the proposed algorithm and K-Means is plotted in Figure 6 for 200 images in dataset. The difference between the number of nuclei segmented by the proposed algorithm and K-Means is also plotted in Figure 6 in order to illustrate variation. However, the results for 40 images in dataset are not plotted in Figure 6 since K-Means produces over-segmented or under-segmented results for these images. Why K-Means produces such results and how the proposed algorithm eliminates these errors are described in discussion section. The comparison results also demonstrate that the deviation between the proposed algorithm and the expert is 2%, whereas the deviation between K-Means and Expert is 5%.

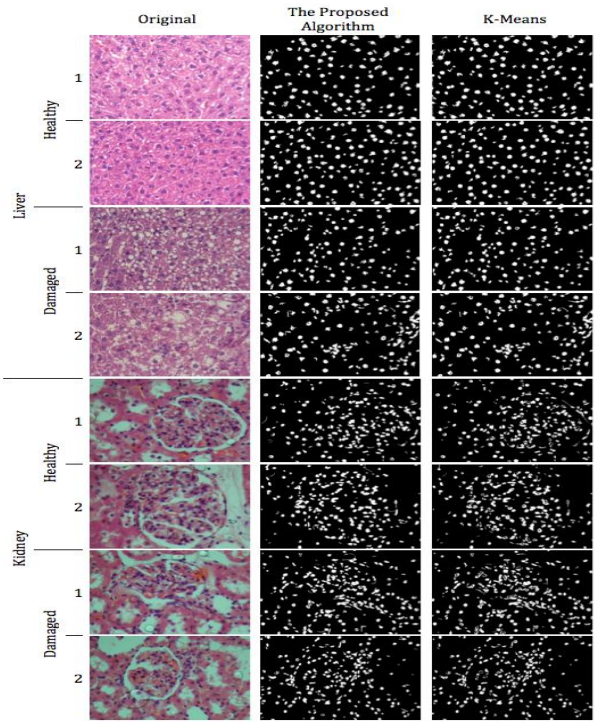

Figure 5. Histopathological liver and kidney tissue images and segmentation results

Table 2. The number of nuclei for images in Figure 5

Kind		No	K-Means	The Proposed Algorithm	Expert
Liver	Healthy	1	148	128	119
		2	161	167	164
	Damaged	1	145	143	143
		2	135	139	139
Kidney	Healthy	1	201	199	196
		2	181	181	176
	Damaged	1	215	211	206
		2	206	217	217

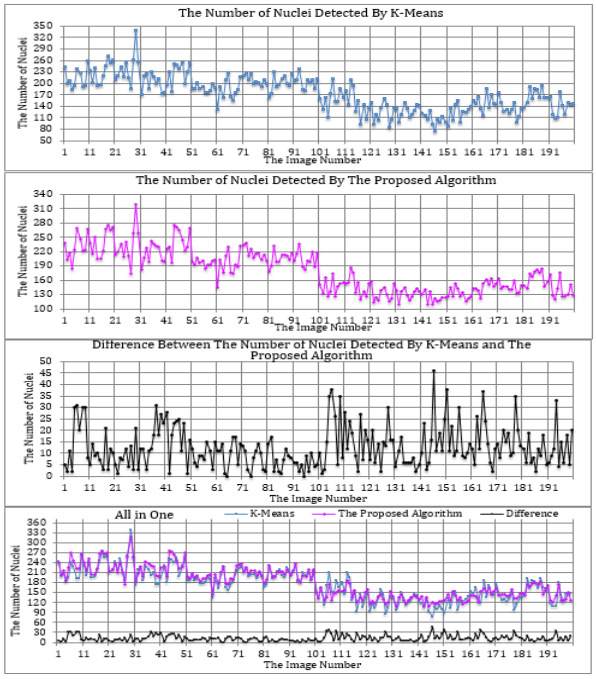

Figure 6. The number of nuclei segmented by the proposed algorithm and K-Means, and difference between them. 1-50: healthy kidney, 51-100: damaged kidney, 101-150 healthy liver and 151:200 damaged liver tissue images.

Table 3. Processing time of the algorithms

Method		Tissue	Kind	Time (second)
K-Means	Segmentation	Liver	Healthy	8.121
			Damaged	9.038
		Kidney	Healthy	26.888
			Damaged	39.315
	Distinguishing	Liver	Healthy	0.126
			Damaged	0.116
		Kidney	Healthy	0.335
			Damaged	0.568
The proposed algorithm		Liver	Healthy	1.350
			Damaged	1.335
		Kidney	Healthy	5.104
			Damaged	8.462

Table 4. Ratio of processing time for the algorithms

Tissue	Kind	Time (second)		
		K-Means (K)	The Proposed Algorithm (P)	Ratio (K/P)
Liver	Healthy	8.247	1.350	6.109
	Damaged	9.153	1.335	6.857
Kidney	Healthy	27.223	5.104	5.334
	Damaged	39.883	8.462	4.713
Average		22.443	3.930	5.575

The segmentation results of two algorithms are generally close to each other except images that K-Means fails to segment. However, there is relatively more difference between some results especially in the liver images as seen in Figure 6, since the difference between the distributions ratios of the components in the liver images, which are stained with different color, is greater than the difference between the distributions ratios of the components in the kidney images. The details are in discussion section.

The proposed method is much more efficient than K-means in terms of processing time. The processing times of K-Means and the proposed algorithm for all images in dataset are shown in Table 3 and Table 4.

The proposed distinguishing algorithm takes much less time than segmentation in K-Means as seen in Table 3. The ratios of total processing time for two algorithms are presented in Table 4 for more convenient comparison of the processing times. Distinguishing and segmentation times of K-Means are summed in Table 4.

The processing time of K-Means changes depending on selection of the initial points for clusters, whereas processing time of the proposed algorithm is constant for an image. The proposed algorithm is approximately 6 times faster than K-Means in average as seen in Table 4.

4.3. Discussion

In histopathological image analysis, experts and CAD systems rely on histological tissue specimens. The complex three-dimensional (3D) specimens consisting of structures such as cell nuclei or cytoplasm is sectioned with a microtome into thin sections, and two-dimensional (2D) images are obtained from these sections. The size, shape, color specification and frequency of nuclei in the sections depend on a variety of parameters, including the orientation and position of the sectioning plane. The nuclei in sections are shown as blue spots in the images. The distance between microtome and nuclei affects the blue color specification of the nuclei in sectioning process. The nuclei appear in dark blue color specification as in Figure 7(b), if the microtome passes through the nuclei. However, the nuclei appear in light blue color specification as in Figure 7(c-e), if the microtome passes over the nuclei. The expert, K-Means, and the proposed algorithm can see and segment dark nuclei easily, whereas determination of the nuclei shown in Figure 7(d-e) is more difficult. In case of many nuclei with light blue color specification, the result of expert and the proposed algorithm is close to each other than K-means as seen in Figure 5 and Table 2.

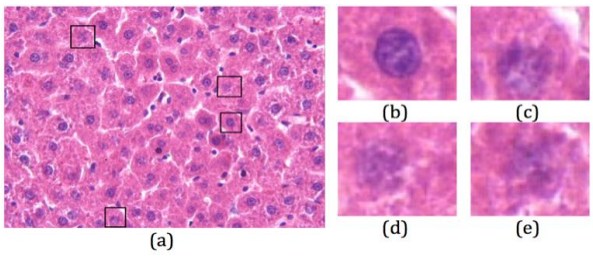

Figure 7. Representation of nuclei in different color specification of blue. (a) A histopathological tissue image (b) A nucleus of (a) in dark blue color specification. (c-e) A nucleus of (a) in light blue color specification.

Imperfections in routine technical procedures in preparation process of the tissue samples may yield more different color specification than three main color specification in H&E stained tissue images. An example of this can be seen in Figure 8(a), where two samples of unexpected color specification are shown in rectangles. These imperfections may hamper performance of segmentation algorithms, while may have no effect on the expert's evaluation. A test image with imperfections shown in Figure 8(a) is segmented with the proposed algorithm and K-Means and evaluated by the expert. While the number of nuclei counted by the expert is 130 and the number of nuclei segmented by the proposed algorithm is 125 for Figure 8(b), K-Means produces different results depending on the cluster initial point selected randomly. There are two different results of K-means as seen in Figure 8(c-d) when the number of clusters is determined as 5. The number of segmented nuclei in Figure 8(d) is 122 that is close to expert and the proposed algorithm results, while there is under-segmentation in Figure 8(c). When the number of clusters is determined as 3, K-Means produces over-segmented results as seen in Figure 8(g). In addition, the number of clusters is increased and selected as 6, K-Means produces under-segmented result as shown in Figure 8(h). The similar results as seen in Figure 8(e-f) are obtained when the number of clusters is selected as 4. The proposed algorithm has no requirements of factors such as number of clusters or initial points, and thus segments images with imperfections accurately.

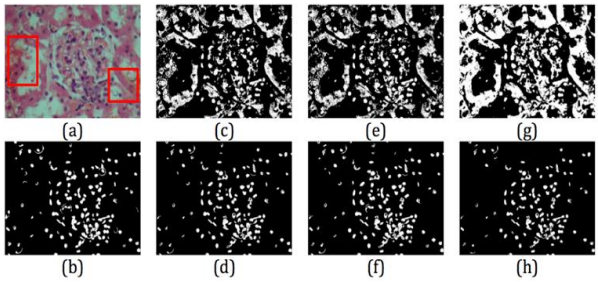

Figure 8. Over and under segmentation example for K-Means

5. CONCLUSIONS

Segmentation is a crucial initial stage of many CAD systems, and the quality of the segmentation algorithm significantly affects the success of the CAD system. In this paper, a novel segmentation algorithm named Probabilistic Nuclei Segmentation Algorithm is proposed to segment H&E stained histopathological images without any requirement of pre-processing, post-processing, and any parameter or a threshold value determined manually. Experimental results show that the proposed algorithm segments histopathological images more accurately and approximately 6 times quickly than K-Means. Additionally, the proposed algorithm is able to yield a success result in case of processing images with imperfections, where K-Means fails to produce success result. The comparison results of the number of nuclei also demonstrate that the deviation between the proposed algorithm and the expert is 2%, while the deviation between K-Means and Expert is 5%.

In addition, the study has two other contributions. First, the distinguishing nuclei cluster algorithm is developed to distinguish nuclei cluster from other clusters for K-means. Second, false positive removal algorithm is proposed to eliminate tiny points of nuclei cluster, which can cause deceptive nuclei evaluation.

REFERENCES

[1] Mills, S.E., "***Histology for Pathologists***", Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, (2012).

[2] Suvarna, K.S., Layton, C. and Bancroft, J.D., "***Bancroft's Theory and Practice of Histological Techniques***", Elsevier Health Sciences UK, (2012).

[3] He, L., Long, L.R., Antani, S. and Thoma, G., "Computer assisted diagnosis in histopathology", ***Seq. Genome Anal. Methods Appl.***, 271–287, (2010).

[4] Fox, H., "Is H&E morphology coming to an end?" ***J. Clin. Pathol.***, 53: 38–40, (2000).

[5] Murphy, D.B. and Davidson, M.W., "***Fundamentals of Light Microscopy and Electronic Imaging***", Wiley-Blackwell, Hoboken, N.J., (2012)

[6] Thomas, G.D., Dixon, M.F., Smeeton, N.C. and Williams, N.S., "Observer variation in the histological grading of rectal carcinoma", ***J. Clin. Pathol.***, 36: 385–91, (1983).

[7] Metter, G.E., Nathwani, B.N., Burke, J.S., Winberg, C.D., Mann, R.B., Barcos, M., Kjeldsberg, C.R., Whitcomb, C.C., Dixon, D.O. and Miller, T.P., "Morphological subclassification of follicular lymphoma: variability of diagnoses among hematopathologists, a collaborative study between the Repository Center and Pathology Panel for Lymphoma Clinical Studies", ***J. Clin. Oncol.***, 3: 25–38, (1985).

[8] Dick, F., VanLier, S., Banks, P., Frizzera, G., Witrak, G., Gibson, R., Everett, G., Schuman, L., Isacson, P., O'Conor, G., Cantor, K., Blattner, W. and Blair, A., "Use of the Working Formulation for Non-Hodgkin's Lymphoma in Epidemiologic Studies: Agreement Between Reported Diagnoses and a Panel of Experienced Pathologists", ***J. Natl. Cancer Inst.***, 78: 1137–44, (1987).

[9] Chan, W.C., Armitage, J.O., Gascoyne, R., Connors, J., Close, P., Jacobs, P., Norton, A., Lister, T.A., Pedrinis, E., Cavalli, F. and others, "A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma", ***Blood***, 89: 3909–3918, (1997).

[10] Serin, F., Ertürkler, M. and Gül, M., "K-nearest unrepeatable cell graph model of histopathological tissue image", ***2015 23nd Signal Processing and Communications Applications Conference (SIU)***, 2585–8, (2015).

[11] Serin, F., Erturkler, M. and Gul, M., "A novel overlapped nuclei splitting algorithm for histopathological images", ***Comput. Methods Programs Biomed.***, 151: 57–70, (2017).

[12] Gunduz, C., Yener, B. and Gultekin, S.H., "The cell graphs of cancer", ***Bioinformatics***, 20: i145–51, (2004).

[13] Ng, H.P., Ong, S.H., Foong, K.W.C., Goh, P.S. and Nowinski, W.L., "Medical image segmentation using K-means clustering and improved watershed algorithm", ***2006 IEEE Southwest Symposium on Image Analysis and Interpretation***, 61–65, (2006).

[14] Petushi, S., Garcia, F.U., Haber, M.M., Katsinis, C. and Tozeren, A., "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer", ***BMC Med. Imaging***, 6: 14, (2006).

[15] Bilgin, C., Demir, C., Nagi, C. and Yener, B., "Cell-Graph Mining for Breast Tissue Modeling and Classification", ***2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society***, 5311–4, (2007).

[16] Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M. and Yener, B., "Histopathological image analysis: A review", ***Biomed. Eng. IEEE Rev. In***, 2: 147–171, (2009).

[17] Kothari, S., Chaudry, Q. and Wang, M.D., "Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques", ***2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE***, 795–8, (2009).

[18] Bilgin, C.C., Bullough, P., Plopper, G.E. and Yener, B., "ECM-aware cell-graph mining for bone tissue modeling and classification", ***Data Min. Knowl. Discov.***, 20: 416–438, (2010).

[19] Malu, G., Balakrishnan, K. and Bodhey, N.K., "Area and volume calculation of necrotic tissue regions of heart using interpolation", ***2011 International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT)***, 728–30, (2011).

[20] Baykara, M., Erturkler, M., Gul, M. and Harputluoglu, M., "Karaciğer Dokusundaki Nekroz Alanın Doku Tabanlı Bölütleme Kullanılarak Belirlenmesi ve Nicemlenmesi", ***Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU)***, Trabzon/Turkey, (2012).

[21] Ozseven, T., Erturkler, M., Nurmuhammed, M., Gul, M. and Harputluoglu, M., "Quantifying the necrotic areas on liver tissues using support vector machine (SVM) algorithm and Gabor filters", ***2012 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)***, 1–5, (2012).

[22] Serin, F., Erturkler, M., Gul, M. and Yigitcan, B., "Non-Alkolik Yağlı Karaciğer Hastalığında Karaciğerdeki Yağ Vakuolleri Oranının Hesaplanması", ***Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU)***, 306–10, (2012).

[23] Serin, F., Erturkler, M., Gul, M. and Yigitcan, B., "Investigating the effects of melatonin and resveratrol agents on non-alcoholic fatty liver disease", ***Glob. J. Technol.***, 3, (2013).

[24] Skodras, A., Giannarou, S., Fenwick, M., Franks, S., Stark, J. and Hardy, K., "Object recognition in the ovary: Quantification of oocytes from microscopic images", ***2009 16th International Conference on Digital Signal Processing***, 1–6, (2009).

[25] Chang, W.-Y., Huang, A., Yang, C.-Y., Lee, C.-H., Chen, Y.-C., Wu, T.-Y. and Chen, G.-S., "Computer-Aided Diagnosis of Skin Lesions Using Conventional Digital Photography: A Reliability and Feasibility Study", ***PLOS ONE***, 8: e76212, (2013).

[26] Veta, M., Pluim, J.P.W., van Diest, P.J. and Viergever, M.A., "Breast Cancer Histopathology Image Analysis: A Review", ***IEEE Trans. Biomed. Eng.***, 61: 1400–11, (2014).

[27] Wang, S., Burtt, K., Turkbey, B., Choyke, P., Summers, R.M., Wang, S., Burtt, K., Turkbey, B., Choyke, P. and Summers, R.M., "Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research, Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research", ***BioMed Res. Int. BioMed Res. Int.***, 2014, 2014: e789561, (2014).

[28] Firmino, M., Morais, A.H., Mendoça, R.M., Dantas, M., Hekis, H. and Valentim, R., "Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects", ***Biomed Eng Online***, 13: 1–16, (2014).

[29] Otsu, N., "A threshold selection method from gray-level histograms", ***Automatica***, 11: 23–27, (1975).

[30] Adams, R. and Bischof, L., "Seeded region growing", ***Pattern Anal. Mach. Intell. IEEE Trans. On***, 16: 641–647, (1994).

[31] Patil, D.D. and Deore, S.G., "Medical image segmentation: a review", ***Int. J. Comput. Sci. Mob. Comput.***, 2: 22–27, (2013).

[32] Zhang, C., Xiao, X., Li, X., Chen, Y.-J., Zhen, W., Chang, J., Zheng, C. and Liu, Z., "White Blood Cell Segmentation by Color-Space-Based K-Means Clustering", ***Sensors***, 14: 16128–47, (2014).

[33] Zhang, D.-Q. and Chen, S.-C., "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation", ***Artif. Intell. Med.***, 32: 37–50, (2004).

[34] Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J. and Chen, T.-J., "Fuzzy c-means clustering with spatial information for image segmentation", ***Comput. Med. Imaging Graph.***, 30: 9–15, (2006).

[35] Kong, H., Belkacem-Boussaid, K. and Gurcan, M., "Cell nuclei segmentation for histopathological image analysis", ***SPIE Medical Imaging***, International Society for Optics and Photonics, 79622R–79622R, (2011).

[36] Zhang, X., Xing, F., Su, H., Yang, L. and Zhang, S., "High-throughput histopathological image analysis via robust cell segmentation and hashing", ***Med. Image Anal.***, 26: 306–15, (2015).

[37] Xu, Y., Zhu, J.-Y., Chang, E.I.-C., Lai, M. and Tu, Z., "Weakly supervised histopathology cancer image segmentation and classification", ***Med. Image Anal.***, 18: 591–604, (2014).

[38] Wienert, S., Heim, D., Saeger, K., Stenzinger, A., Beil, M., Hufnagl, P., Dietel, M., Denkert, C. and Klauschen, F., "Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach", ***Sci. Rep.***, 2: 503, (2012).

[39] Al-Kofahi, Y., Lassoued, W., Lee, W. and Roysam, B., "Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images", ***IEEE Trans. Biomed. Eng.***, 57: 841–52, (2010).

[40] Kecheril, S.S., Venkataraman, D., Suganthi, J. and Sujathan, K., "Automated lung cancer detection by the analysis of glandular cells in sputum cytology images using scale space features", ***Signal Image Video Process.***, 9: 851–63, (2013).

[41] Kothari, S., Phan, J.H., Stokes, T.H. and Wang, M.D., "Pathology imaging informatics for quantitative analysis of whole-slide images", ***J. Am. Med. Inform. Assoc.***, 20: 1099–108, (2013).

[42] Ray, S. and Turi, R.H., "Determination of number of clusters in k-means clustering and application in colour image segmentation", ***Proceedings of the 4th international conference on advances in pattern recognition and digital techniques***, Calcutta, India, 137–143, (1999).

[43] *https://www.mathworks.com/help/images/color-based-segmentation-using-k-means-clustering.html*, "Color-Based Segmentation Using K-Means Clustering - MATLAB & Simulink Example", (2017).

[44] He, L., Chao, Y. and Suzuki, K., "A Run-Based Two-Scan Labeling Algorithm", ***IEEE Trans. Image Process.***, 17: 749–56, (2008).