

PAPER • OPEN ACCESS

An Automated Segmentation and Counting of Ki67 Cells in Meningioma Using K-Means Clustering Technique

To cite this article: Fahmi Akmal Dzulkifli *et al* 2019 *J. Phys.: Conf. Ser.* **1372** 012060

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

An Automated Segmentation and Counting of Ki67 Cells in Meningioma Using K-Means Clustering Technique

Fahmi Akmal Dzulkifli*¹, Mohd Yusoff Mashor¹, and Hasnan Jaafar²

¹School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600, Arau, Perlis, Malaysia.

²Department of Pathology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, 16150, Kubang Kerian, Kelantan, Malaysia

*fahmiakmaldzulkifli@gmail.com

Abstract. Meningioma is a type of primary brain tumours. The meningiomas account for about one-third of all primary brain tumours. Image segmentation plays an important role in image analysis, especially detecting the tumours or cancerous areas in medical images. The output images from the segmentation prominently affect the system in detecting the tumour cells. Currently, the pathologists use the ‘eye-balling’ estimation technique to count the Ki67 cells. This technique was known as a time-saving measure. However, it has poor reliability and accuracy in counting the Ki67 cells. This paper proposed an automatic Ki67 cell counting in meningioma by using k-means clustering approach. The k-means clustering was used to segment the Ki67 cells and then the cells were classified into positive and negative Ki67 cells. The proposed system has been tested on 12 histopathological meningioma images. The proposed system is compared to the manually segmented images that have been validated in prior by the pathologists. The results show that the proposed system was able to segment the Ki67 cells with an average accuracy of 95.29%. The sensitivity and specificity of the proposed system were also high with an average of 93.56% and 97.39%, respectively.

1. Introduction

A brain tumour is referring to a collection of abnormal cells, which often forms a mass of tissue inside the brain. Generally, the brain tumour is divided into two categories, which are primary and metastatic brain tumour [1]. Meningioma is a type of primary brain tumour that arises in the three thin layers of tissue called meninges. According to CBTRUS statistical report, meningioma was the most frequently reported histology among the primary brain tumours, which accounts for 37.1% [2]. Based on the World Health Organization (WHO) grading scale, the meningioma is usually benign and grows. Due to this fact, these tumours cells are difficult to identify in some locations unless the size of the tumour is very large to be discovered. Ki67 is a nuclear antigen which response to a monoclonal antibody, MIB-1. Ki67 is known as an independent prognostic factor for survival rates, which consists of all stages and grade categories [3]. The Ki67 labelling index (LI) is defined as the percentage of immunoreactive tumour cell nuclei. The Ki67 LI is one of the measurement technique used to measure the normal and abnormal proliferation. Jonat and Arnold [4] classified the Ki67 LI into three stages, namely low (LI < 15%), intermediate (16%-30%), and highly proliferating (LI > 30%).

There are various image segmentation techniques were reported in the previous studies such as thresholding, wavelet transform, fuzzy c-means, quick shift technique, and watershed model [5]. K-means clustering is an unsupervised clustering algorithm used for dividing the images into few regions



based on image colour property [6]. The clustering can be classified into two types, which are Hard Clustering (Exclusive Clustering) and Soft Clustering (Overlapping Clustering) [7]. For a hard clustering method, each data point belongs to a single cluster and the properties of this element cannot distribute to another cluster [7]. In soft clustering, each data point can belong to more than one cluster. In this study, K-means clustering was applied to partition the histopathological image into two groups namely the positive and negative Ki67 cells. This paper is organized as follows: Section 2 reviews previous studies used k-means clustering technique in medical imaging. Section 3 discusses the procedures and methodology for segmenting the Ki67 cells. Section 4 presents the experimental results obtained from the proposed technique. Section 5 concludes the paper.

2. K-Means Clustering

Nowadays, the use of image segmentation especially in medical fields is more focusing on segmenting tissues and organs. The application also includes tumour detection, surgical planning, brain development study, heart segmentation, and analysis of the cardiac images. A proper segmentation will simplify analysis to yield a more accurate result.

Sharma *et al.* [8] listed the problems existing in computed tomography (CT) and MR images. These include the partial volume effect, different type of artefacts, and noise due to sensors and related electronic system. Thus, the application of image segmentation will be useful in segmenting of an image. Over the years, many approaches have been proposed to segment medical images image by using k-means clustering technique.

Karmilasari *et al.* [9] developed a system that can determine the stage of breast cancer by using k-means clustering in the mammogram images. Three categories of mammogram images were used in the study which is dense-glandular, fatty, and fatty-glandular. At the beginning stage, the system will determine the region of interest (ROI). The proposed system cropped the input images to reduce the size of the input images. Otsu's thresholding was applied to separate the background and foreground of the images. Next, the system segmented the images by using a region growing technique. Then, the calculation of the suspected area size was performed based on the segmented images. K-means clustering was used to partition the suspected area into three stages (I, II, and III). Based on the results, the system can determine the stage of breast cancer based on the calculation of the area of the suspected object.

Bombale and Patil [10] proposed an algorithm to detect lung nodule from CT images by using k-means clustering. The segmentation was done to extract the part of the lungs from the images for detecting the probable cancerous area. Four types of nodules were selected in the study which are Well-circumscribed, Juxta-vascular, Pleural Tail, and Juxta-pleural. The classification process was to determine whether the nodule is benign or malignant. The classification was based on shape and size, appearance and growth rate. Based on the result, the proposed system was able to detect 494 samples from 525 samples of chest CT images. The proposed system obtained a good detection with an overall accuracy of 94.09%. This paper applied the k-means clustering technique to partition the positive and negative Ki67 cells to calculate the Ki67 LI.

3. Automated Ki67 Counting

In this study, the proposed automated Ki67 counting was operated on histopathological images of meningioma.

3.1. Image Acquisition

A total of 12 histopathological images were captured at the Department of Pathology, Hospital Universiti Sains Malaysia (HUSM). In this study, the positive Ki67 cells were stained using diaminobenzidine (DAB), while the haematoxylin was used to stain the negative nuclei. These images were captured under 40x magnification using an Olympus BX51 microscope and Cell[^]F software that works as an interface to the digital camera that attached with the microscope. The resolution of the captured images were 1360×1024 pixels and the images were saved in (*.jpg) format.

3.2. Image Enhancement

In this study, the image enhancement is essential for increasing the contrast and brightness of the Ki67 cells. In contrast enhancement process, the Ki67 images were enhanced by using Contrast-Limited Adaptive Histogram Equalization (CLAHE). To increase the contrast of an object, the original image must be converted into a colour space that has the image luminosity components. In the beginning, the Ki67 images were converted from RGB to L*a*b* colour space. The RGB colour space needs to be converted into CIE XYZ colour space first and followed by the L*a*b* colour space. The conversion from RGB colour space into CIE XYZ colour space can be calculated as follows [11]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Next is the conversion from CIE XYZ to L*a*b* colour space by using the equations below [12]:

$$L = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \quad (2)$$

$$a^* = 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \quad (3)$$

$$b^* = 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right] \quad (4)$$

where X_n , Y_n , and Z_n are the tristimulus value of the reference white. X , Y , and Z describe as any colour that can be perceived by an average human observer. Next step was the luminosity value selection. The range of the values needs to be scaled into [0 1] since the luminosity component image was type double. For the best contrast enhancement result, the luminosity value was selected at 0.87. This value was the minimum value that has been selected manually based on analysis of 12 images. Then, the CLAHE technique is applied to the L component and the results of the L image was multiplied with the luminosity value. The output image from the contrast enhancement process was then retrieved back into RGB colour space. For increasing the brightness of the image, a constant value was selected and added to each of the pixels in the image. After performing analyses using all the captured images, it has been found that the best value was 20 since this value produced the best segmentation results. If the constant value is higher or lower than 20, it will degrade the segmentation results.

3.3. Image Segmentation

In this study, the k-means clustering technique was used to classify between the positive and negative Ki67 cells. At first, the proposed system will remove the background of the output Ki67 image from the enhancement technique. The background removal was done using Otsu's thresholding method. The purpose of this step is to remove the background of that image so that only the objects are remained. Then the resultant image was converted back to the original image.

At this moment, there were two visible objects left in the image which are the positive and negative Ki67 cells. It is known that the positive Ki67 cells appeared to be in granular brown colour while the negative Ki67 cells appeared in blue colour. These two colours were useful in k-means clustering, where the number of clusters needs to be specified at first before the segmentation process. Next step was to fill in the current image with 'dummy' pixels. The reason for adding the 'dummy' pixels is to control any input images that execute in the proposed system so that the system able to separate the objects correctly based on the specified clusters. The 'dummy' pixel contained the colour information of positive Ki67 cells. In this study, the 'dummy' pixel was set to $R = 170$, $G = 130$, and $B = 60$ that represents the brown colour. Before proceeding further process, the proposed system will add a 'dummy' pixel for every 100 pixels to the image. Due to the presence of the 'dummy' pixels inside the image, it certainly does not disturb the k-means clustering process since these pixels can be removed later using image filtering method. Afterward was the implementation of k-means clustering technique to segment the Ki67 cells.

At first, the captured image was converted into $L^*a^*b^*$ colour space using Equation (1), (2), (3), and (4). Next was to choose k points into the space represented by the objects that are being clustered. For this study the number of clusters was $k = 2$. This is because the current image (output image after the background removal) only has two colours which are brown and blue. Then was the calculation of the Euclidean distance, d by using the equation [6]

$$d = \|x_i - c_i\| \quad (5)$$

where x_i was the pixel in an image and the c_i was initial group centroids. Then was to assign each pixel to the group that has the closest centroid. When all objects have been assigned, recalculate the positions of the k centroids. The process will be repeated until the centroids are no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated. Next step was to label every pixel in the image using the results from k-means clustering. 'Dummy' pixels were added by using the RGB value that mention earlier in the previous paragraph. After adding the 'dummy' pixels, the process was continued to classify the cluster 1 for positive Ki67 cells and cluster 2 for negative Ki67 cells. Lastly, the noises and the 'dummy' pixels were removed using area opening method.

3.4. Feature Extraction

In this study, there were three main types of features that were selected based on the features of Ki67, namely circularities, area, and solidity based features. The area was expressed by the area of fully segmented Ki67 cells which includes the positive and negative Ki67 cells. The circularities features used to differentiate between the Ki67 cells with unwanted objects like artefacts. Solidity was used to measure the density of the cell. These three features will be the inputs for the proposed algorithm to determine whether the segmented objects were categorized in Ki67 cells or unwanted objects. Those objects will be discarded from the resultant images when it does not meet the requirement of the range value. The extracted image of positive and negative Ki67 cells was later saved into a bitmap (*.bmp) file extension.

3.5. Counting Ki67 Cells

In this step, the binary image from the feature extraction process will be used to count the Ki67 cells. Currently, the binary image only has two possible values (0 or 1) for each pixel. The '0' represented as the background while the '1' referred to Ki67 cells. The Sobel filter was applied to find the edges of the objects. After that, each detected object was labelled with different numbers from 1 to n . This can be performed by calculating the centre of gravity (centroid) to determine the position of the segmented cells and label them.

3.6. Calculation of Ki67 LI

Equation (6) below presents the formula to calculate the Ki67 LI. The equation was used after obtaining the number of positive and negative Ki67 cells. By following the WHO Classification of Tumours of the Central Nervous System, the tumours was labelled as grade I (LI < 7.99%), grade II (8% to 19.99%), and grade III (LI > 20.00%) [13].

$$\text{Ki67 Index} = \frac{\text{No. of positive Ki67 Cells}}{\text{No. positive Ki67 Cells} + \text{No. of negative Ki67 Cells}} \times 100\% \quad (6)$$

4. Results And Discussion

The proposed system was developed using MATLAB software version 2018. A personal computer which runs on an Intel Core i7-5500U, 2.4 GHz, processor with 16.0 GB RAM that operates on Microsoft Windows 10 Pro was used to develop the system. In order to test the performance of the segmentation method, the proposed segmentation algorithm was compared with manual segmentation. The manual segmentation was carried out using an open source image analysis software known as ImageJ. This software was used to segment the Ki67 cells manually and the resultant images were validated by HUSM pathologists. Quantitative analysis was performed to measure the performance of

the proposed system in segmenting Ki67 cells. The evaluation of the performance depended on the calculation of segmentation accuracy, sensitivity, and specificity. These analyses are calculated by:

$$Accuracy = \left(\frac{TP+TN}{(TP+TN+FP+FN)} \right) \times 100\% \quad (7)$$

$$Sensitivity = \left(\frac{TP}{(TP+FN)} \right) \times 100\% \quad (8)$$

$$Specificity = \left(\frac{TN}{(TN+FP)} \right) \times 100\% \quad (9)$$

where TP was defined as the number of pixels that were correctly segmented as positive KI67 cells. FP indicated as the number of pixels that incorrectly segmented as positive Ki67 cells. TN was referred to the number of pixels that were correctly segmented as a negative Ki67 cell, while FN signified the number of pixels that incorrectly segmented as negative Ki67 cells. Table 1 shows the segmentation results of the proposed system.

Table 1. Segmentation results of the proposed system

Input Image	Accuracy (%)	Sensitivity (%)	Specificity (%)
Cell 01.bmp	97.08	95.41	98.88
Cell 02.bmp	96.88	95.75	98.07
Cell 03.bmp	97.32	96.55	98.11
Cell 04.bmp	95.94	92.75	99.64
Cell 05.bmp	94.73	90.77	99.53
Cell 06.bmp	94.26	90.08	99.42
Cell 07.bmp	95.49	96.10	94.90
Cell 08.bmp	95.20	95.99	94.44
Cell 09.bmp	94.35	96.56	92.34
Cell 10.bmp	94.47	91.21	98.29
Cell 11.bmp	93.39	89.61	97.95
Cell 12.bmp	94.36	91.91	97.11
Average (%)	95.29	93.56	97.39

Based on Table 1, the results showed that the proposed system was able to segment the Ki67 images with an average accuracy of 95.29%. All of the images shown to have good accuracy with more than 90%. The results also showed that the proposed system able to detect the Ki67 cells with an average sensitivity value of 93.56%. Next, Table 2 shows the comparison between the proposed automated counting system with the manual counting system for positive and negative Ki67 cells. The table also showed the comparison between the Ki67 index between the proposed counting system with manual counting system. The manual counting was calculated by the pathologists.

Table 2. Comparison between the proposed automated counting with manual counting

Input Images	Manual Counting			Proposed Automated Counting System		
	Immunopositive Ki67	Immunonegative Ki67	Ki67 LI (%)	Immunopositive Ki67	Immunonegative Ki67	Ki67 LI (%)
Cell01.bmp	51	407	11.14	61	365	14.32
Cell02.bmp	62	340	15.42	67	328	16.96
Cell03.bmp	70	351	16.63	78	292	21.08
Cell04.bmp	21	305	6.44	18	392	4.39
Cell05.bmp	25	312	7.42	17	411	3.97
Cell06.bmp	35	373	8.58	28	509	5.21
Cell07.bmp	41	316	11.48	82	241	25.39
Cell08.bmp	37	295	11.14	50	265	15.87
Cell09.bmp	48	287	14.33	62	266	18.90
Cell 10.bmp	79	345	18.63	78	461	14.47
Cell 11.bmp	97	361	13.32	76	415	15.48
Cell 12.bmp	168	308	35.29	133	265	33.42

Based on Table 2, it showed that the proposed system by using the k-means clustering technique was able to calculate Ki67 cells. From this table also, the proposed system showed a good result in counting positive Ki67 cells. For negative Ki67 cells, there was a large gap of counting between the proposed automated counting system and the manual counting. The existing problem happened may be due to the automated number of clusters (k) that was specified earlier in the segmentation process. The average execution time to count Ki67 cells for one image was 17 seconds.

5. Conclusion

This paper proposed an automatic Ki67 cell counting for histopathological meningioma images. The proposed system consists of six steps which are image acquisition, image enhancement, image segmentation, feature extraction, counting Ki67 cells, and calculation of Ki67 LI. The Ki67 images were enhanced using the CLAHE technique and followed by the segmentation process using k-means clustering. Three features were selected to extract the Ki67 cells, which includes of circularity, area, and solidity. A quantitative analysis was used to measure the performance of the proposed system. Based on the result, the proposed system was capable to segment and count the Ki67 cells. The proposed system was able to obtain an excellent segmentation result with an average accuracy of 95.29%. The proposed system also showed a high sensitivity and specificity values with an average of 93.56% and 97.39 & respectively. For future developments, the proposed system may be remodelled to improve the accuracy of the algorithm especially in detecting the Ki67 cells.

Acknowledgments

The author would like to thank the pathologists from Department of Pathology, University Science of Malaysia (HUSM) for helping and contributing to this study.

References

- [1] Agravat R R and Raval M S 2016 Brain Tumor Segmentation Towards a Better Life *CSI Commun.* **40** 31–35
- [2] Ostrom Q T, Gittleman H, Truitt G, Boscia A, Kruchko C and Barnholtz-Sloan J S 2018 CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015 *Neuro. Oncol.* **20** 1–86
- [3] Li L T, Jiang G, Chen Q and Zheng J N 2015 Ki67 is a promising molecular target in the diagnosis of cancer (Review) *Mol. Med. Rep.* **11** 3 1566–1572
- [4] Jonat W and Arnold N 2011 Is the Ki-67 labelling index ready for clinical use? *Ann. Oncol.* **22** 3 500–502
- [5] Suganya and Menaka 2014 Various Segmentation Techniques in Image Processing: A Survey *Int. J. Innov. Res. Comput. Commun. Eng.* **2** 1 1048–1052
- [6] MacQueen J 1967 Some Methods for Classification and Analysis of Multivariate Observations *5th Berkeley Symposium on Mathematical Statistics and Probability 1967*.
- [7] Bora D J, Gupta A K and Khan F A 2015 Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation *Int. J. Emerg. Technol. Adv. Eng.* **5**,2 192–203
- [8] Sharma N and Aggarwal L M 2010 Automated Medical Image Segmentation Techniques *J. Med. Phys.* **35** 1 3–14
- [9] Karmilasari, Hanum Y, Widodo S, Hermita M, Putri N and ETP L 2014 Sample K-Means Clustering Method for Determining the Stage of Breast Cancer Malignancy Based on Cancer Size on Mammogram Image Basis *Int. J. Adv. Comput. Sci. Appl.* **5** 3 86–90
- [10] Bombale A and Patil C G 2017 Segmentation of Lung Nodule in Ct Data Using K-Mean *Int. J. Electr. Electron. Data Commun.* **5** 2 36–39
- [11] Kaur A and Kranthi B 2012 Comparison between YCbCr Color Space and CIE Lab Color Space for Skin Color Segmentation *Int. J. Appl. Inf. Syst.* **3** 4 30–33
- [12] Tkaleie M and Tasie J F 2003 Colour spaces - Perceptual, Historical and Applicational Background *IEEE Reg. 8 EUROCON 2003. Comput. as a Tool.* **2** 304–308

- [13] Louis D N, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee, W K, Ohgaki H, Wiestler O D, Kleihues P, and Ellison D W 2016 The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary *Acta Neuropathol.* **131** 6 803–820