



# Journal of Medical Sciences

ISSN 1682-4474

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

*JMS (ISSN 1682-4474) is an international, peer-reviewed scientific journal that publishes original article in experimental & clinical medicine and related disciplines such as molecular biology, biochemistry, genetics, biophysics, bio-and medical technology. JMS is issued eight times per year on paper and in electronic format.*

*For further information about this article or if you need reprints, please contact:*

Yazan M. Alomari,  
Pattern Recognition Research  
Group,  
Center for Artificial Intelligence  
Technology,  
Faculty of Information Science  
and Technology,  
University Kebangsaan Malaysia,  
43600, Bangi, Malaysia

## Random Patch Probabilistic Density Algorithm for Tissue Localization from the Whole Slide Images

<sup>1</sup>Yazan M. Alomari, <sup>1</sup>Siti Norul Huda Sheikh Abdullah,  
<sup>2</sup>Reena Rahayu Md Zin and <sup>1</sup>Khairuddin Omar

Objects localization from the whole slide images is one main issues that can be handled by image processing techniques which is important for both, the medical and computer science fields. In this study, random patch probabilistic density method is proposed for localization the tissue from the whole slide histology images. The proposed method is a simple localization method, it based on foreground density feature inside a virtual box and the box represents a randomly selected region from the foreground of the image. The proposed method then used to localize the tissue from the whole slide image which represents the Region of Interest (ROI) in this case. In medical imaging, in the process of the analyzing the whole slide tissue microscopic images which is considered an extremely important step in histopathological image analysis and diagnosing, this analysis includes localizing the tissue region from the whole slide in a bounding box, before the scanning process starting. The proposed method able to localize the objects adaptively without predetermining the number of objects or clusters to be found as in K-means and Fuzzy C-means. After that, the proposed method, results, evaluation and comparison are explained.

**Key words:** Localization, whole slide tissue, possible objects, density feature, ROI, tissue

## INTRODUCTION

The problem of localizing the ROI from images can be addressed as clustering problem in image processing field (Gurcan *et al.*, 2009; Hiary *et al.*, 2013; Lopez *et al.*, 2012; Sertel, 2010). In medical domain, when deals with whole slide tissue images, according to the clinical setting (Hiary *et al.*, 2013; Alomari *et al.*, 2009), pathologists need to manually determine the tissue region in a bounding box, as a pre-step to enter this tissue slide to scanner.

Current clustering algorithms such as K-means and Fuzzy C-means can handle the localization challenge. However, they still have limitations of computation time needed and pre-knowledge of the number of clusters. If they used a clustering hybridization method, the regions found vary in size and the number of regions found is a small number by Lopez *et al.* (2012), because the regions found was based on density of required objects which depends on the connectivity between them. Based on that, a method that localizes unknown number of regions and adaptive localization method was suggested to deal with more challenges than previous methods.

In the evolution of digitizing the whole slide tissue, many challenges appear that requires the pathologist to localize the ROI manually. When the whole slide enter to the digital pathology scanner without pre-localize the tissue from the whole slide, this leads to consume a huge memory and produce a huge image size (Hiary *et al.*,

2013). Thus, the technician manually fixes a box around the required tissue to be scanned. The purpose of this step is approximately determining the area to be scanned by the digital pathology scanner from the whole slide. The scanner will only scan the area inside the box. This box must include all the significant tissue regions from the whole slide. Figure 1 shows some tissue samples localized in a bounding box from the whole slide, these samples different in tissue size, location, shape and color. Automating the process of localizing of the tissue from the whole slide in a bounding box can saves the pathologists time, efforts and reduces the memory and the produced image size. This localization method creates bounding box around the tissue and if there are some noise or small tissue pieces scattered in the slide which are insignificant to pathologists, it will ignore them.

This task is considered as a pre-processing step before entering the slide to the high-resolution digital slide scanner. This step is to determine the area from the slide to be scanned. In the result, this step reduces the memory, image size and the scanning time. In fact, it reduces between 60 and 80% from the final image size and up to 40% from the scanning time.

However, manually localize the tissue from the whole slide by pathologists is an extremely tedious, subjective and time-consuming task. An accurate segmentation and localization mechanism that localizes the tissue from the whole slide may help pathologists and save their time.

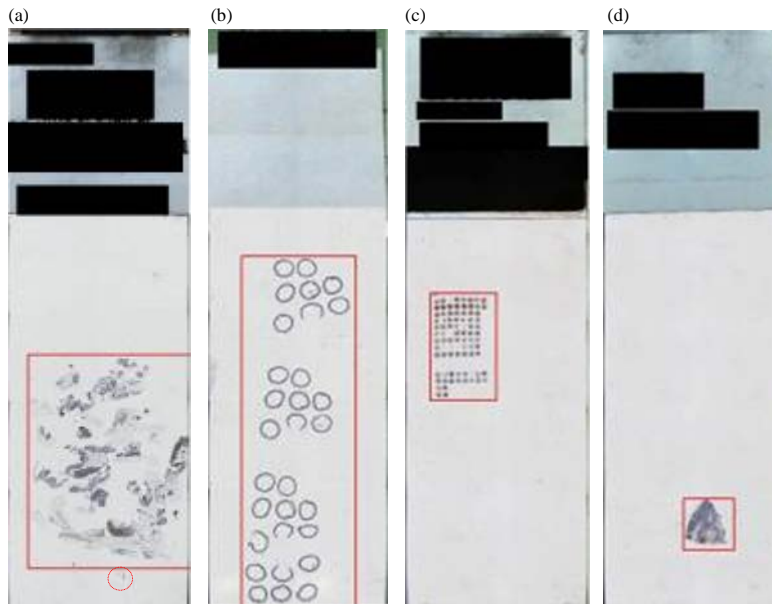


Fig. 1(a-d): Sample cases for the localized tissue from the whole slide in a bounding box, (a) Medium tissue size bounded in a box and red circle represents an ignored small-scattered tissue piece, (b) Big size tissue, (c) Small tissue size and (d) Very small tissue size

Using of clustering and localization methods is highly depending on the application, imaging modalities and other factors. Each image modality has its own characteristics to be extracted to perform the segmentation. Therefore, there is no segmentation method which has acceptable results for all imaging modalities (Lopez *et al.*, 2012). According to that challenge, medical imaging segmentation is still known as a problem for this field (Hiary *et al.*, 2013). Different approaches in medical image segmentation and localization, some approaches based on heuristics, region growing, edge detection and thresholding methods (Alomari *et al.*, 2014). Other approaches use machine learning techniques, including supervised, unsupervised and semi supervised (Alomari *et al.*, 2009; Hiary *et al.*, 2013). While other approaches need prior information in the forms of atlases (Cuadra *et al.*, 2004).

Many researchers have investigated the whole slide tissue analysis using unsupervised techniques, Nadernejad and Sharifzadeh (2013) presented a new pixonal-based method for image segmentation. They form a pixonal image using bilateral filter which is used as a kernel function. Some advantages achieved by using this filter such as, decreasing image noise and helping in smoothing the image too. In addition, avoiding over segmentation problem and removing unwanted noises from the environment. In the next step, fuzzy C-means is used in order to segment the obtained pixonal image. Their experimental results show less computational and better accuracy compared to other segmentation techniques.

Li (2008) presents some modifications in the original K-means clustering algorithm. He claims the original K-means have some limitations such as low efficiency in the way that K-means perform the centroid calculations which affect on the K-means efficiency. Therefore, a new method proposed to overcome this limitation. Additional preprocessing steps are added to the input images to speed up and enhance the clustering process. Then, he improves the clustering seed method through moving the seeds clustering toward intensive data areas. In addition, original K-means in some cases stuck in the local maximum solution. Solving this problem is done through using the proposed method as local searching process in order to inlay in local search structure of the iteration. The proposed method outperforms the original K-means method through the local search and extends the searching range. A comparison with original K-means using 3D medical volume images is conducted; his proposed method achieves 94.7% accuracy while original K-means accuracy is 77.5%. Other important limitation in K-means which are not discussed in their study, is pre-determining the number of clusters.

Some researchers use unsupervised learning methods to make segmentation and localization the tissue from the whole slide images. Hiary *et al.* (2013) present a method to segment the tissue from the whole slide image and localize it in a bounding box using K-means algorithm. They aim to extract only the region of interests which represented her in the tissue, bounded in the box, before entering this slide to the slide scanner. This step saves both the scanning time and memory space required. Unsupervised clustering K-means method achieves 96% accuracy. Using unsupervised clustering methods still needs to pre-determine the number clustering. On the other hand, it does not need a ground truth in their case, Unlike their previous work (Alomari *et al.*, 2009), when they use supervised machine learning method to make the same task of segmentation and localization. They make a hybrid method to combined heuristic methods with parametric machine learning. Color, intensity, texture and spatial distribution features were used in segmentation and localization the tissue. In addition to the use of principal component analysis in feature reduction, then training in two layers in back propagation neural network. The accuracy achieved is 96% with ground truth and data training required.

Few studies Lopez *et al.* (2012) and Elie *et al.* (2003), are conducted in the literature that localizing the focus-point regions from Ki-67 whole tumor slide images. Hybridization in clustering methods are done by Lopez *et al.* (2012), they present a hybrid method to localize hot spots in the whole slide tissue for ki-67 stained histology images. Hot spot regions in the whole-slide are the process of identifying the tumor regions that exhibit high proliferating activity. They propose a clustering method which is able to localize an unknown number of clusters. This is done by proposing hybridization between Hierarchical Clustering (HC) and nonhierarchical clustering techniques. They propose a hybrid method combined between DBSCAN algorithm and the standard single linkage (HC) method. DBSCAN algorithm is used to initiate the single linkage method by specific instances, instead of starting from the individual data. They apply their method on Ki-67 stained whole slide pathology images; compare the proposed method results with the manual hot spots labeled by pathologist. Results showed some improvements in consistency between pathologists in the hot spots task. Furthermore, support pathologists in quantitative descriptors. The resulted clusters from this hybrid clustering method used in this study can be highly variable in terms of shape, size and density. Unfortunately, variability in cluster size may confuse pathologists, when going for higher magnifications to this region for further analysis. Therefore, if all hot spots clusters are bounded by a box

with same size for all boxes and found all regions suggested to be visited by pathologist, this will make easier to the pathologists to visit all boxes for further analysis if needed. In addition, this method is highly sensitive to parameters which is hard to determine for all cases.

Elie *et al.* (2003) present a simple method for quantifying the focus point regions that represents the stained cells from the whole slide tissue. Two thresholding steps then used, first is to extract the tissue, the second is to extract stained cells regions from the tissue. After that, a morphology close operator used to combine all neighbored pixels. Next, reconstruction of the size 10 is done using the open morphology operator, to keep only large stained areas. At the end and logical operator applied between a manual marked mask image and the binary image, stained pixels and focus point regions. The drawback of this technique is that some numerous parameters are not easy to determine and not easy to set as a fix value to be suitable for all cases, such as thresholding, because they use fixed thresholding in some steps. Thus, these values can significantly influence the results.

Our objective in this study is to develop and validate an algorithm that localizes the tissue in a bounding box.

**MATERIALS AND METHODS**

Current clustering algorithms such as K-means and Fuzzy C-means can handle the localization challenge. However, they still have limitations of computation time needed and pre-knowledge of the number of clusters (Li, 2008; Nadermejad and Sharifzadeh, 2013). These limitations are solved in Random Patch Probabilistic Density (RPPD) proposed method. Therefore, a new technique for localization method aims to localize all ROI from image. This technique localizes all regions based on

density feature. It is an adaptive localization technique which able to identify all ROI from the image without pre-knowledge about the number of clusters, as in K-means and Fuzzy C-means. Therefore, Random Patch Probabilistic Density (RPPD) method was proposed to identify all ROI from the image.

The RPPD proposed method for localization the ROI is a simple localization method, it based on foreground density feature inside a virtual box and the box represents a randomly selected region from the foreground of the image. The proposed RPPD method then tested on different types of datasets. It used in localizing the tissue from the whole slide images. In addition, it used in localizing the possible objects from the satellite images for searching the missing MH370 plane. In this localization method, three general steps are done as shown in Fig. 2.

First, remove the whole slide label and extract the glass area that contains the tissue using the crop method. The label region is a standard area for all cases as shown in Fig. 3. The second step is a global image segmentation technique using Otsu thresholding method. This step aims to extract the tissue from the whole slide image. The third step is to follow a local structural segmentation approach on the extracted tissue only, by using the proposed RPPD method. This step aims to identify a particular structure in the tissue. This structure in this case refers to the extracted tissue from the whole slide image. The purpose of using RPPD method on this dataset is to localize the tissue in one bounding box. Thus, after finding all focus-point regions from the whole slide image, minimum and maximum values of (x, y) coordinates was chosen from all found focus-points. Then it draws bounding box that includes the tissue. The proposed RPPD method will consider each focus point region based on the tissue density inside the box which means, it will ignore the small pieces of the tissue scattered in the glass slide.

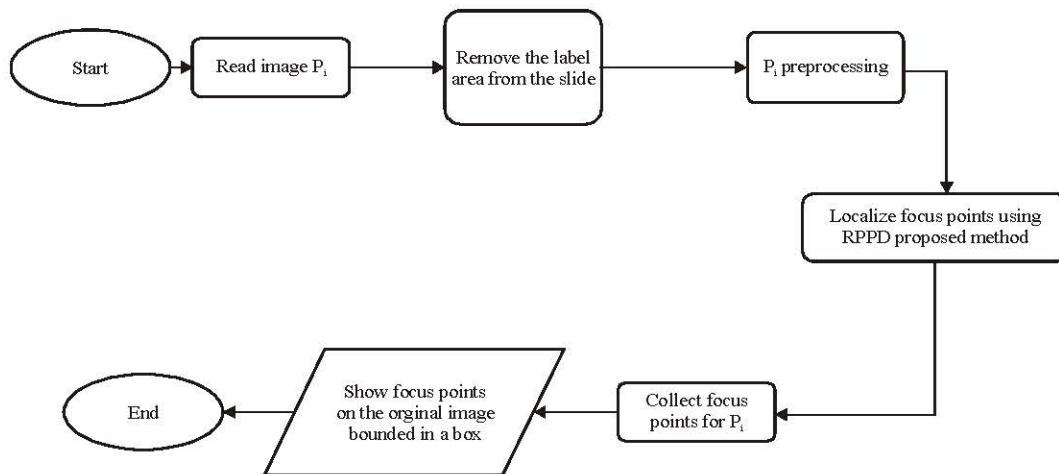


Fig. 2: General methodology for ROI localization proposed method

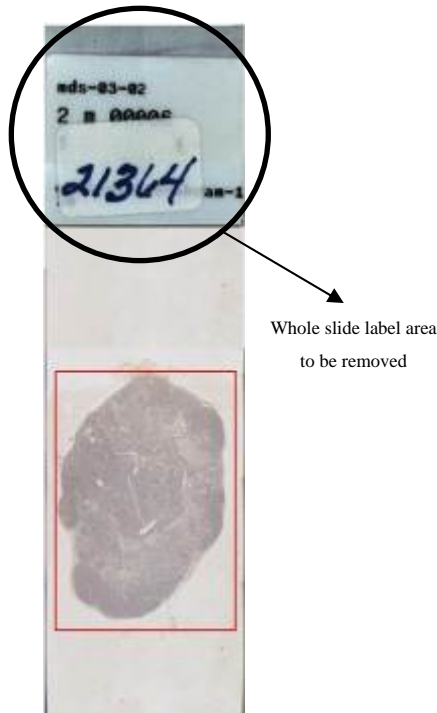


Fig. 3: Whole glass-slide tissue image, the area inside the circle represents the label area to be removed shown in the circle

**Preprocessing:** The proposed RPPD method for ROI localization works with binary images. Whole slide tissue images used in this dataset are colored images. These images captured by a digital camera. Therefore, in such images, some brightness problems appeared while capturing the images. Thus, in this phase, whole slide RGB images follow some preprocessing steps to solve the problem of brightness and to convert the image to gray, then to a binary image. These steps are as follows:

- Step 1:** The RGB  $P_i$  image converted to gray-scale image by eliminating the hue and saturation information while retaining its luminance
- Step 2: Histogram equalization:** This step uses a Contrast-Limited Adaptive Histogram Equalization (CLAHE) method to transform the values of the gray-scale image to enhance the contrast of the image. The CLAHE method works locally in the image instead of the whole image. This carried out by dividing the image into small regions called tiles. Next, histogram equalization is applied for each tile to enhance the contrast and then the results are included in the whole image

**Step 3:** The gray  $P_i$  image is converted into a binary image using the Otsu thresholding method to extract the tissue from the whole-slide image. The Otsu method chooses the threshold to minimize the intra class variance of the black and white pixels

**Proposed Random Patch Probabilistic Density (RPPD):** In general, segmentation in images is defined by the regions to be identified as in Eq. 1.  $FPR_k \Omega$  where,  $\Omega$  is the segmented tissue and need to be segmented into regions  $FPR_k$ :

$$\bigcup_k^N FPR_k = \Omega (\sim FPR), 0 \leq N \leq (m \times n) \quad (1)$$

where,  $N$  is the number of regions and  $R_j \cap R_k = \phi, \forall j \neq k$ ,  $(m \times n)$  is the image size. However, good image segmentation softens the condition of a hard subset to only one region by assigning probabilities for pixels to lie in regions (Alomari *et al.*, 2009).

**Terminology definition I:**

- $P_i$ : Whole slide tissue image
- $\Omega$ : Array of pixels for the segmented tissue in  $P_i$
- $\Omega_k$ : Number of tissue pixels (black pixels) inside the candidate box
- $\Omega_k^c$ : Number of non-tissue pixels (white pixels) inside the candidate box
- $F_i$ : Random pixel selected from  $\Omega$
- CFP: Candidate focus-point region
- $D_i$ : Density feature of the tissue inside candidate CFP
- FPR: CFP becomes a true focus-point region if meets the  $D_i$  feature
- $T_{min}$ : Minimum number of pixels remaining in  $\Omega$  to continue searching for more focus-points
- $T_r$ : Threshold value represents the accepted ratio to considered CFP as FPR

After segmenting the whole slide tissue image using Otsu thresholding method, segmented image mostly represents stained tissue regions which is needed in the case of localizing the stained tissue images. Figure 4a illustrates the original whole slide tissue, Fig. 4b illustrates the original image after the whole slide label was removed, Fig. 4c illustrates the gray scale image and Fig. 4d illustrates the segmented tissue after applying Otsu thresholding. Therefore, in the next step, proposed method should localize all focus-point regions that represent the tissue as shown in Fig. 4e. The final output after using RPPD method on this dataset is to localize the tissue in one bounding box. Thus, after finding all focus point regions from the whole slide image, minimum and maximum values of  $(x, y)$  coordinates was chosen from all found focus points and draw bounding box that contains the tissue as shown in Fig. 4f.

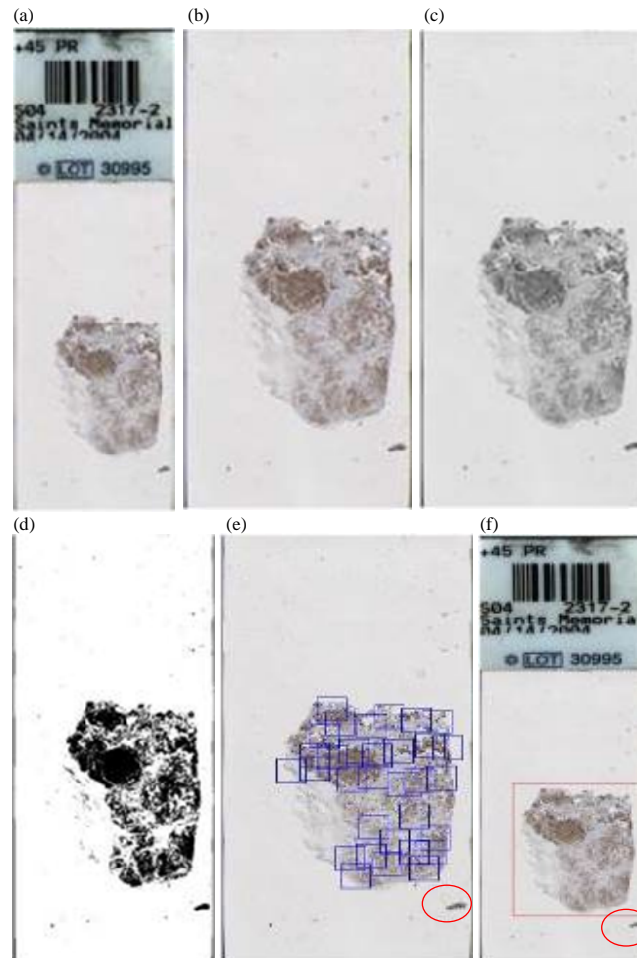


Fig. 4(a-f): (a) Original tissue image, (b) Tissue slide region cropped image, (c) Gray scale image, (d) Segmented tissue image in binary after Otsu thresholding, (e) Tissue after RPPD localizes all focus points, the small tissue piece in the red circle is not considered as a focus point and (f) Final result for RPPD shows the bounding box without including the scattered small piece of tissue

**Basic idea for the proposed RPPD:** The proposed RPPD method involves binary images. The black pixels represent the tissue; therefore, all of the tissue pixels are stored in tissue array  $\Omega$ . A pixel,  $F_i$ , is chosen randomly from  $\Omega$ . A virtual box is then drawn with a size of  $(40 \times 40)$  pixels and  $F_i$  is set to be the center of this box. This box is considered a Candidate Focus-Point region (CFP).

In the next step, this CFP is checked based on density-feature criteria to decide whether it will be considered a CFP. Figure 5 shows the randomly selected  $F_i$  and the virtual box centered by  $F_i$ . The proposed RPPD method solved the main limitation for the current clustering methods which is the pre-knowledge concerning the number of clusters. In addition, all tissue parts are checked locally from the image.

Many experiments were performed to determine the choice of box size. The trade-off between larger and smaller box size is the number of boxes at the end and the final localization accuracy. Each true box represents a focus point region. When the size of the box is small, the number of focus point regions will be large in the outcome. If the size of the box is large, the number of focus point regions will be small in the outcome. In addition, the bigger image size requires bigger box size to achieve higher percentage of accuracy.

In the case of localizing the tissue from the whole slide, RPPD focuses on detecting the tissue and tissue borders accurately, because the final output should have the minimum bounding box which contains the tissue. Thus,  $(40 \times 40)$  pixels are the box size was selected for this experiment.

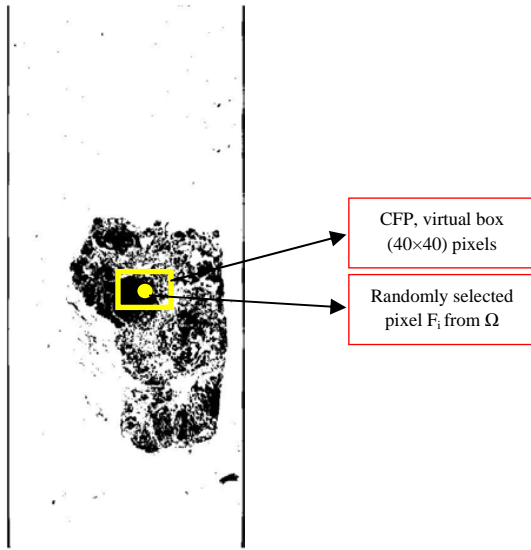


Fig. 5: Randomly selected point from tissue and CFP

**Determining candidate focus point region:** In the next step, RPPD identifies the density feature of the tissue inside the box based on Eq. 2 as shown in Fig. 5. The RPPD accepts or rejects this region based on the equation shown in Eq. 3. If the value of  $D_i$  is greater than a threshold value  $T_r$ , then CFP is considered a true FPR. Next, RPPD removes all pixels inside the box from  $\Omega$  and moves on to select a new  $F_i$ . Otherwise, if the value of  $D_i$  is less than a threshold value  $T_r$ , RPPD rejects this CFP and restarts by selecting a new  $F_i$ :

$$D_i = \frac{\Omega^t}{\Omega^t + \Omega^f} \quad (2)$$

where, box size (40x40):

$$\sum \text{CFP} = \begin{cases} \text{Reject} & D_i < T_r \\ \text{Accept} & D_i \geq T_r \\ \text{Remove } \Omega^t \text{ from } \Omega & D_i \leq 5\% \end{cases} \quad (3)$$

If density inside the box is less than 10%, RPPD removes the pixels inside this box. In some cases, during the slide preparation process, some small pieces of tissue spread through the slide. These tissue pieces are not significant regions for the pathologists. Therefore, RPPD ignores such pieces of tissue as shown in Fig. 6.

**RPPD iterations number:** The proposed RPPD method continues searching for focus-point regions in the image  $P_i$  until meets one of two conditions:

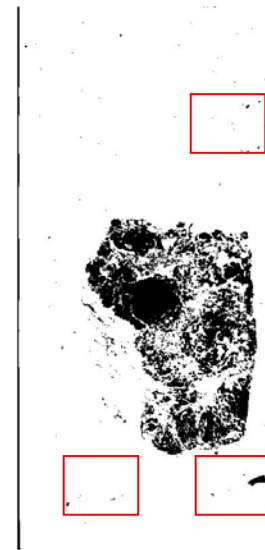


Fig. 6: Insignificant regions of tissue found and ignored by RPPD showed inside the red boxes, thus in the final bounding box it will not included

- The number of pixels in  $\Omega$  is less than  $T_{min}$
- The number of iterations exceeds 50 times without identifying any FPR

The value of  $T_{min}$  is technically defined by the maximum number of tissue pixels inside the box required to keep this CFP region rejected from proposed RPPD method to consider as a FPR. It can be formulated as the Eq. 4:

$$T_{min} = \frac{\Omega^t}{\Omega^t + \Omega^f} \quad (4)$$

where,  $T_{min} < 5\%$ .

This number of remaining tissue pixels in  $\Omega$  is not sufficient to decide a FPR even if these pixels found in a candidate box. Therefore, if the number of remaining tissue pixels has reached  $T_{min}$ , RPPD proposed method terminates the iterations.

Conversely, a number of iterations exceeding 50 without identifying any FPR, indicates that most of the FPRs are found and the probability of finding a new FPR is very weak in the image  $P_i$ . This value of 50 is determined based on experimental tests and because the image size is not large.

**Accept CFP as a Focus-Point Region (FPR):** In this step, if the  $D_i$  value inside the box is greater than a threshold value  $T_r$ , then the RPPD method deems this CFP as FPR.



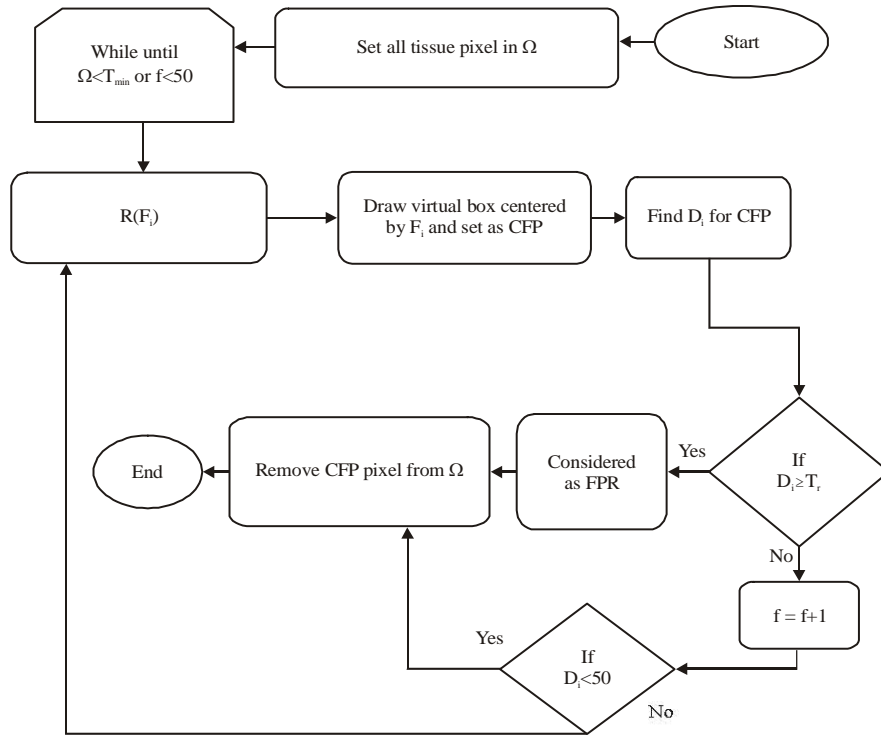


Fig. 7: Proposed RPPD method workflow

If density inside the box is less than 5%, the proposed RPPD method removes the pixels inside this box. In some cases, during the slide preparation process, some small pieces of tissue spread through the slide. These tissue pieces are not significant regions for the pathologists. Therefore, the proposed RPPD method ignores such pieces of tissue as shown in Fig. 6. The workflow of the proposed RPPD method for each  $P_i$  is shown in Fig. 7.

**Whole glass slide images dataset:** The problem of localizing the ROI from the whole slide tissue image can be addressed as clustering problem in image processing field (Gurcan *et al.*, 2009; Hiary *et al.*, 2013; Lopez *et al.*, 2012; Sertel, 2010). In this dataset, 246 images used for the whole-tissue slide which were used in (Alomari *et al.*, 2009; Hiary *et al.*, 2013), the proposed RPPD method was used to localize the tissue from the whole slide image and fix a box around the tissue. The required memory and time to scan the whole slide was very large. The purpose of this step was to determine approximately the area to be scanned by the digital pathology scanner from the whole slide. The scanner will only scan the area inside the box. This box represents the minimum area from slide that contains the localized tissue. This method will save scanning time and memory space required.

**Evaluation method:** For the evaluation method used and according to the evaluation method used by Alomari *et al.* (2009) and Hiary *et al.* (2013), if the box localizes all the tissue with minimum non-tissue area, it was considered true localization. If the box missed a significant area of the tissue, it was considered a localization error and the image was counted as an error (localization error). The localization accuracy was calculated as in Eq. 5:

$$\text{Accuracy} = \left(1 - \frac{L_{\text{Error}}}{N}\right) 100\% \quad (5)$$

The box size used in the proposed method was  $(40 \times 40)$  pixels and the  $T_r$  threshold value used was 5%. After identifying all of the focus point regions (boxes) in the image, the maximum box that contains all localized small boxes in the image was then found. This step was performed to localize the tissue in one bounding box.

## RESULTS

For the 246 whole slide tissue images, the localization accuracy was 97.3% using the proposed RPPD method. It outperformed the method used by Alomari *et al.* (2009) and Hiary *et al.* (2013). They used

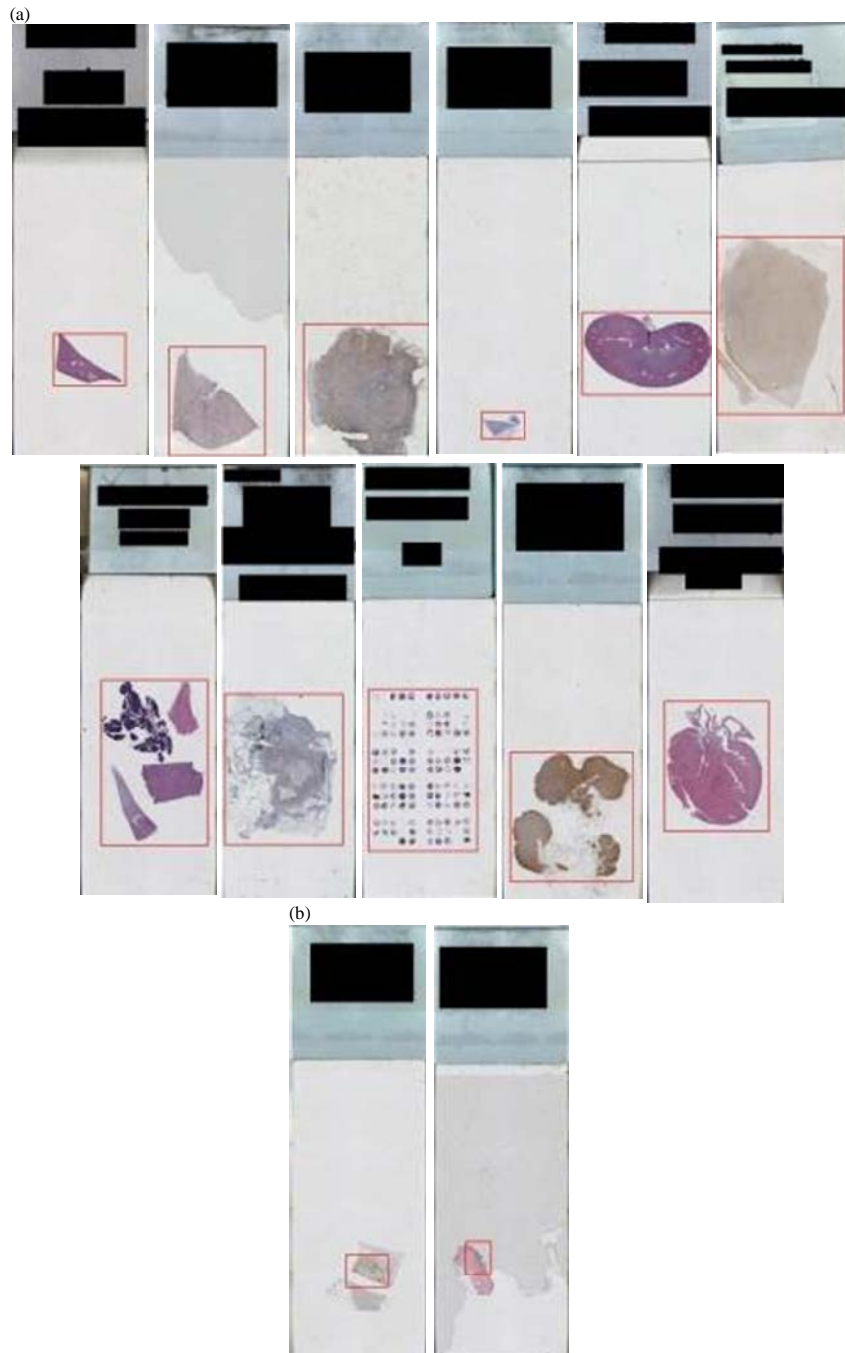


Fig. 8(a-b): (a) Sample correct tissue and (b) Sample incorrect tissue localization results for the proposed RPPD method

unsupervised and supervised learning methods to segment the tissue from the whole slide and then bounded the tissue in a box. They achieved 96% of the localization accuracy using the same dataset.

The proposed RPPD method was able to localize tissue with different tissue size, shape, position and

colors. Figure 8 shows some sample results for the proposed RPPD method when used for the tissue localization from the whole slide; Fig. 8a shows some correct tissue localization bounded in a box and Fig. 8b shows some incorrect tissue localization with significant tissue missing.

Table 1: Summary of tissue localization results for supervised, unsupervised learning methods, K-means, fuzzy C-means clustering methods and the proposed RPPD method

Method	Localization accuracy (%)
Supervised learning method Alomari <i>et al.</i> (2009)	96.0
Unsupervised learning method Hiary <i>et al.</i> (2013)	96.0
K-means clustering method (k = 40)	91.8
Fuzzy C-means clustering method (k = 25 )	73.5
RPPD proposed method	97.3

**Comparison of the results with other methods:** The first comparison was with unsupervised learning method (Hiary *et al.*, 2013) and supervised learning method (Alomari *et al.*, 2009), their accuracy was high because it was based on set of different type of features extracted from the tissue. The advantage of using unsupervised learning methods in that case was it does not require training. Both methods, supervised and unsupervised achieved 96% accuracy of the final tissue localization and bound it in a box.

However, localization problem can be handled by known clustering methods such as K-means and Fuzzy C-means. Thus, two more experiments were conducted for the tissue localization in a bounding box, using the same dataset. The experiments were conducted using both K-means and Fuzzy C-means clustering methods. The experiments were conducted using 40 clusters for K-means and 25 clusters for Fuzzy C-means methods. Both of K-means and Fuzzy C-means are centroid based, thus, their results were approximately equal. Therefore, a different clusters number used for each method in order to discuss the consequence of using different clusters number in the results. This experiment was conducted in order to compare the proposed RPPD method results with other benchmark methods.

Regarding to K-means and Fuzzy C-means, same preprocessing steps were performed to extract the tissue. The binary-extracted tissue pixels which mostly represents the tissue are clustered using K-means (Pham *et al.*, 2005) and Fuzzy C-means (Bezdek *et al.*, 1984) based on the tissue pixel coordinates. The clustering methods clustered the tissue pixels based on distances between pixels, indicating that the tissue pixels are close together in a cluster. On the other words, clustering the tissue is performed based on the tissue pixel concentration or density, similar to that based on RPPD. Unfortunately, both K-means and Fuzzy C-means methods able to localize the tissue, but it also have many found clustered on non-tissue regions. Which causes the resulted bounding box is larger than localized tissue, that is not accepted due it costs more memory and image size resulted from the scanner. The final localization accuracy achieved using both K-means and Fuzzy C-means methods was 91.8 and 73.5%, respectively. Table 1 summarizes all localization accuracy results for all methods discussed above.

## DISCUSSION

In the proposed RPPD, the small boxes that represents the focus point regions, mostly determines the tissue borders. Then, one bounding box is created to contain all focus point regions identified. When dealing with such case, accurately localize the tissue in one bounding box is essential, because extra non-tissue areas included with the bounding box, means extra size in the scanned output image. Thus, small box size with less density ratio will increase the accuracy of localization process. Figure 9a shows different cases of tissue in the slide and focus point regions localized using the proposed RPPD method. Furthermore, insignificant regions will be ignored and only the significant regions from the tissue will localized as in Fig. 9c. After identifying the focus point regions, the bounding box drawn to contains all located focus point regions based on its coordinates as shown in Fig. 10a-c.

Regarding to K-means clustering methods, it performs well when using the number of clusters is 40. When the tissue size is big, the clustered focus point regions were mostly from the tissue with some clusters found on non-tissue regions as in Fig. 9d. In case of small tissue size, it performance was also not accurate, some of clustered focus point regions were from the tissue and most of them were from non-tissue regions in the image slide Fig. 9e-f. Non-tissue regions can be noise, or the slide glass borders or any small insignificant scattered pieces of the tissue as in Fig. 9f in the red markers. Furthermore, it is still very challenging task to determine the correct number of clusters due to different cases in the dataset, such as size and number of tissue pieces.

However, Fig. 10d-f shows the final bounding box results for the discussed cases. Although the localization accuracy achieved by K-means is 91.8%, but still these final results is not accepted due to very large area of non-tissue regions was included in bounding box.

Regarding to Fuzzy C-means clustering methods, it performs well when using the number of clusters is 25. When the tissue size is big, the clustered focus point regions were mostly from the tissue with few clusters found on non-tissue regions. In addition, this cause to detect the tissue borders not accurately and the bounding box missed some significant areas from the tissue as in Fig. 9g. In case of small tissue size, it performance was also not accurate, some of clustered focus point regions were from the tissue and most of them were from non-tissue regions in the image slide Fig. 9h-i. Furthermore, it is still very challenging task to determine the correct number of clusters due to different cases in the dataset, such as size and number of tissue pieces.

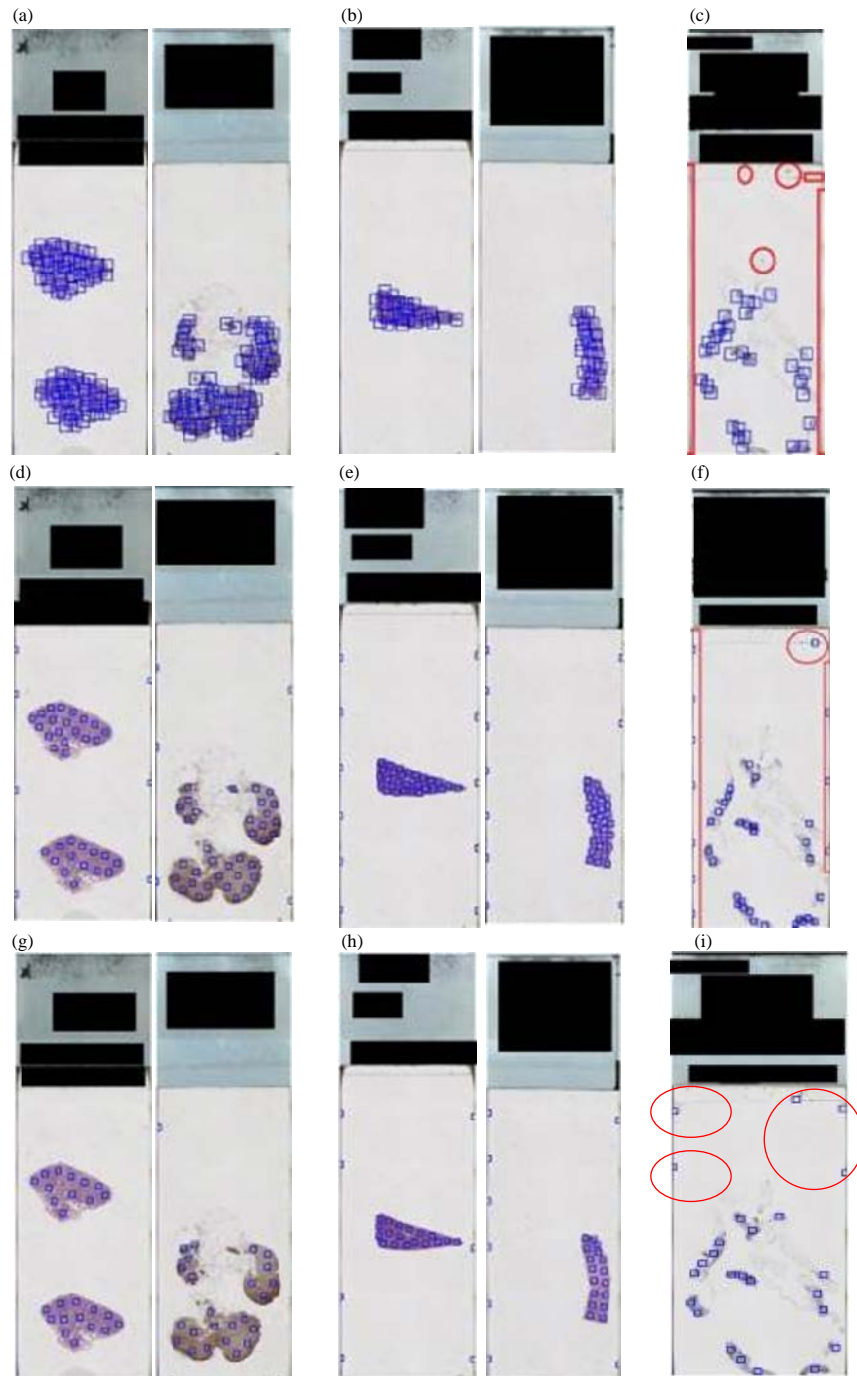


Fig. 9(a-i): (a-c) Sample results for the proposed RPPD shows the focus point regions localized for the whole slide tissue image, (d-f) Sample results for the K-means clustering method with 40 clusters shows the focus point regions localized for the whole slide tissue image, (g-i) Sample results for the Fuzzy C-means clustering method with 25 clusters shows the focus point regions localized for the whole slide tissue image, (a, d, g) Sample of more than big tissue cases, (b, e, h) Sample of small tissue cases and (c, f, i) Sample of small reacted tissue pieces



Fig. 10(a-i): (a-c) Sample of final bounding box result for the proposed RPPD method for the previous mentioned cases, (d-f) Sample of final bounding box result for the K-means clustering method using 40 clusters and (g-i) Sample of final bounding box results for the Fuzzy C-means clustering method using 25 clusters

However, Fig. 10g-i shows the final bounding box results for the discussed cases. Although the localization accuracy achieved by Fuzzy C-means is 73.5% but still these final results is not accepted due to very large area of non-tissue regions was included in bounding box.

## CONCLUSION

Random patch probabilistic density method was proposed and discussed in this study, the proposed method based on density feature for localizing objects. The aim of this localization method is to handle the localization challenges and drawbacks that previous methods were suffering. The proposed method consists from two major steps; partition the whole image into smaller partitions; for better performance when dealing with big size images. Then randomly check the chosen box density to decide to accept or reject this box region. The proposed method tested using different types of datasets. It used to localize the tissue in a bounding box from the whole slide images; the achieved localization accuracy was 97.3% which considered better than achieved accuracy by the previous methods.

In addition, the proposed method outperforms K-means and Fuzzy C-means clustering methods in localization accuracy in the tissue localization dataset. In addition, unlike K-means and Fuzzy C-means clustering methods, RPPD adaptively find all required clusters without pre-determining the number of clusters. The factors and equations used for this method have been explained, in addition to presents the results.

## ACKNOWLEDGMENTS

This study would like to thank to University Kebangsaan Malaysia and Ministry of Education for providing facilities and financial support under “AP-2012-019 Automated Medical Imaging Diagnostic Based on Four Critical Diseases: Brain, Breast, Prostate and Lung Cancer” and FRGS/1/2014/ICT07/UKM/02/5 grant entitled “Overlap irregular shape descriptor based on non-linear approach for pathology images”. Special thanks to: Dr. Reena Rahayu Md Zin who had worked hard in giving ideas.

## REFERENCES

Alomari, R.S., R. Allen, B. Sabata and V. Chaudhary, 2009. Localization of tissues in high-resolution digital anatomic pathology images. Proceedings of the SPIE Medical Imaging Conference, February 8-12, 2009, Lake Buena Vista, FL., USA.

- Alomari, Y.M., S.N.H.S. Abdullah, R.Z. Azma and K. Omar, 2014. Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm. *Comput. Math. Methods Med.* 10.1155/2014/979302
- Bezdek, J.C., R. Ehrlich and W. Full, 1984. FCM: The fuzzy  $c$ -means clustering algorithm. *Comput. Geosci.*, 10: 191-203.
- Cuadra, M.B., C. Pollo, A. Bardera, O. Cuisenaire, J.G. Villemure and J.P. Thiran, 2004. Atlas-based segmentation of pathological MR brain images using a model of lesion growth. *Trans. Med. Imaging*, 23: 1301-1314.
- Elie, N., B. Plancoulaine, J.P. Signolle and P. Herlin, 2003. A simple way of quantifying immunostained cell nuclei on the whole histologic section. *Cytometry Part A*, 56: 37-45.
- Gurcan, M.N., L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot and B. Yener, 2009. Histopathological image analysis: A review. *Biomed. Eng.*, 2: 147-171.
- Hiary, H., R.S. Alomari and V. Chaudhary, 2013. Segmentation and localisation of whole slide images using unsupervised learning. *IET Image Process.*, 7: 464-471.
- Li, X., 2008. A volume segmentation algorithm for medical image based on K-means clustering. Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, August 15-17, 2008, Harbin, China, pp: 881-884.
- Lopez, X.M., O. Debeir, C. Maris, S. Rorive, I. Roland, M. Saerens and C. Decaestecker, 2012. Clustering methods applied in the detection of Ki67 hot-spots in whole tumor slide images: An efficient way to characterize heterogeneous tissue-based biomarkers. *Cytometry Part A*, 81: 765-775.
- Nadernejad, E. and S. Sharifzadeh, 2013. A new method for image segmentation based on Fuzzy C-means algorithm on pixonal images formed by bilateral filtering. *Signal Image Video Process.*, 7: 855-863.
- Pham, D.T., S.S. Dimov and C.D. Nguyen, 2005. Selection of K in K-means clustering. *Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci.*, 219: 103-119.
- Sertel, O., 2010. Image analysis for computer-aided histopathology. Ph.D. Thesis, The Ohio State University, Ohio.