Segmentation of Tumour Region on Breast Histopathology Images for Assessment of Glandular Formation in Breast Cancer Grading

P A S Nor Rahim, N Mustafa\(^1\)*, H Yazid\(^1\), T Xiao Jian\(^2\), S Daud\(^3\) and K S A Rahman\(^4\)

\(^1\)Biomedical Electronic Engineering Department, Faculty of Electronic Engineering Technology, University Malaysia Perlis, 02600 Arau, Perlis.
\(^2\)Department of Electrical and Electronic Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College (TARUC), Kuala Lumpur.
\(^3\)Faculty of Chemical Engineering Technology, University Malaysia Perlis.
\(^4\)Hospital Tuanku Fauziah, 01000 Kangar, Perlis.

*nazahah@unimap.edu.my

Abstract. Breast cancer is the most silent killer among cancers nowadays. NHG system is widely accepted worldwide as a gold standard in providing the overall grade to breast cancer. One of the breast cancer features used in the NHG system is tubule formation. Assessment of tubule formation requires pathologist to identify tumour regions. However, colour variation on breast histopathology could influence tumour regions detection on breast histopathology images. Manual identification of tumour regions using microscope may also vary between pathologists. Thus, automatic segmentation is crucial to segment tumour regions. In this study, a simple approach of segmentation was proposed to segment tumour region on breast histopathology images. The proposed segmentation involved three stages: pre-processing, segmentation and post-processing. The proposed approach using GHE and median filter in the pre-processing stage; Otsu thresholding in the segmentation stage and; morphological operation and pixel removal in the post-processing stage was found able to segment the tumour region with average segmentation accuracy of 90.4 %.

1. Introduction
Breast cancer is the most silent killer among cancers nowadays. Malaysian National Cancer Registry Report (MNCR) 2007-2011 reported that the breast cancer is on the first ranking of common cancers with 18,343 numbers of cancer survival (17.7%) which occurred among female residents in Malaysia [1]. Breast cancer occurs when cells divide and grow uncontrollably in the breast. Cancerous cells proliferate the healthy tissue and destroy them. The breast cancer could be detected if the patient undergoes some cancer diagnosis such as mammogram or ultrasound [2][3]. Detection of breast cancer can also be done using biopsy. In biopsy, breast cells and tissues in the specimens are examined under a microscope. In the case of breast cancer, a standard grading procedure known as Nottingham
Histopathological Grading (NHG) system is used to grade the breast cancer. NHG system provides scores for three breast cancer features. The features are glandular or tubule formation, nuclear pleomorphism and mitotic activity [4]. The score for each feature is ranging from 1 to 3, where score 3 is for the most abnormal condition. Total score from these three features provide grade for the breast cancer.

Tumour regions provide meaningful information especially in glandular formation assessment [5]. However, colour variation on breast histopathology could influence tumour regions detection on breast histopathology images. Manual identification of tumour regions using microscope may also vary between pathologists. Recently, breast tumour regions segmentation on histopathological images has received major attention from researchers and exhibits tremendous growth. Several approaches such as pixel-wise [6], clustering [7], and neural network [8] were used to achieve a promising tumour region segmentation on breast histopathological images. In this study, a simple approach of image processing techniques was employed to segment tumour region on breast histopathology images.

2. Methods

The image processing techniques used in this study were divided into three stages: pre-processing, segmentation and post-processing. Phyton programming (PyCharm Community Edition 2020) was used to perform the image processing tasks.

2.1 Data
The breast histopathological slides were obtained from the Pathology Department, Hospital Tuanku Fauziah, Kangar, Perlis, Malaysia. Hematoxylin and Eosin (H&E) stain was performed on the tissue slides. Then the slides were scanned using Aperio CS2 WSI image scanner to convert the slides into digital form (digital images) and in tiff format.

2.2 Pre-processing
Image enhancement and filtering techniques were used in the pre-processing stage. The purpose of image enhancement is to improve the image quality [9][10]. For this purpose, global histogram equalization (GHE) and contrast-limited adaptive histogram equalization (CLAHE) methods were used. GHE spreads the most repeated values of intensity by extending the image’s intensity range whereas CLAHE enhances the contrast of the input image by changing the hue and saturation value [11]. For filtering, median filter was used.

2.3 Image segmentation
The pre-processed images were then applied with segmentation techniques to demarcate between tumour and non-tumour regions. For this purpose, two segmentation algorithms were chosen: Otsu thresholding and k-means clustering. Otsu thresholding technique selects an optimal threshold value by maximizing the between class variance of the gray levels in the foreground and background [12]. While k-means algorithm goal is to minimize the sum of squared distances for each data point to its cluster centroid in all clusters [13].

2.4 Post-processing
Sharpness, colour balance, dust spots, and other image flaws can be improved with post processing [14]. In this stage, morphological operations and pixel removal were applied on the segmented breast cancer histopathology images. Morphology refers to a group of image processing operations that work with images based on their shapes. Each pixel in the image is adjusted based on the value of other pixels in its vicinity in a morphological operation. The opening operation was implemented on the segmented breast histopathology images then followed with closing operation. Lastly, small unwanted regions were removed from the images.
2.5 Analysis of segmentation techniques
The performance between GHE and CLAHE in enhancing image contrast can be assessed using peak signal to noise ratio (PSNR). Accuracy was used to evaluate the segmentation performance between Otsu thresholding and k-means clustering.

3. Results and Discussion
The proposed segmentation method was implemented on 10 breast histopathology images. Figure 1 shows three samples of the breast histopathology images named as Sample 1 (a), Sample 2 (b) and Sample 3 (c) in grey scale colour. In the pre-processing stage, the images underwent contrast enhancement technique using GHE and CLAHE. The results for GHE and CLAHE are shown in Figures 2 and 3 respectively. The enhanced images were then filtered using median filter. The results of contrast enhanced and filtered images are shown in Figures 4 and 5 respectively. The results of pre-processing were compared by calculating PSNR. It was found that the combination of GHE and median filter provides higher PSNR value of 28.37 dB. Thus, the resultant images of GHE and median filter were opted for further segmentation process.

![Figure 1. Original breast histopathology images in grey scale. (a) Sample 1, (b) Sample 2 (c) Sample 3](image)

![Figure 2. Results of contrast enhancement using GHE](image)

![Figure 3. Results of contrast enhancement using CLAHE](image)
The pre-processed images were then segmented using Otsu thresholding and k-mean clustering. For k-means clustering, the number of centres was set to three. Then, the segmented images of Otsu thresholding and k-means clustering underwent post-processing techniques and the final results of segmentation for Otsu thresholding and k-means clustering are shown in Figures 6 and 7 respectively.

The accuracy of segmentation for both proposed segmentation methods was calculated (Table 1). From Table 1, the average accuracy of Otsu thresholding (i.e., 90.4 %) was found higher than the average accuracy of k-means clustering (i.e., 66.75 %). The finding using Otsu thresholding was also found to provide higher accuracy when comparing with [15].
Table 1. Accuracy of segmentation for Otsu thresholding and k-mean clustering

| Image     | Otsu thresholding | k-means clustering |
|-----------|-------------------|--------------------|
| Sample 1  | 93.19 %           | 82.80 %            |
| Sample 2  | 98.00 %           | 56.27 %            |
| Sample 3  | 93.48 %           | 59.86 %            |
| Sample 4  | 90.36 %           | 67.37 %            |
| Sample 5  | 84.63 %           | 85.88 %            |
| Sample 6  | 88.93 %           | 60.71 %            |
| Sample 7  | 86.08 %           | 39.79 %            |
| Sample 8  | 92.91 %           | 92.35 %            |
| Sample 9  | 90.37 %           | 29.79 %            |
| Sample 10 | 86.06 %           | 92.65 %            |
| Average   | 90.40 %           | 66.75 %            |

The combination of GHE and median filter provides better image quality compared to CLAHE and median filter in the pre-processing stage. For segmentation, Otsu thresholding provides better segmentation compared to k-means clustering by providing higher accuracy of more than 90%. The segmentation using Otsu thresholding was also able to distinguish between tumour and non-tumour regions. The morphological operations (i.e., opening and closing) and pixel removal technique used in the post-processing stage have enhanced the segmentation results by providing cleaner images.

4. Conclusion
In conclusion, Otsu thresholding method was able to segment tumour regions on the breast histopathology images. Implementing GHE and median filter in the pre-processing stage and; morphological operation and pixel removal in the post-processing stage have enhanced the segmentation results. The segmented tumour region in the breast histopathology images perhaps could be used to assist pathologists in the glandular formation assessment for breast cancer grading.

Acknowledgement
The authors gratefully acknowledge the financial support from the Fundamental Research Grant Scheme (FRGC) under a grant number of FRGS/1/2016/SKK06/UNIMAP/02/3 from the Ministry of Higher Education Malaysia. The protocol of this study has been approved by the Medical Research and Committee of National Medical Research Register (NMRR) Malaysia (NMRR-17-281-34236).

References
[1] Azizah Ab M, Nor Saleha IT, Noor Hashimah A, Asmah ZA, Mastulu W 2016 Malaysian Nasional Cancer Registry Report 2006-2011 Nasional Cancer Institute, Malaysia.
[2] Veta M 2013 Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. *PLoS ONE* 8 1-12.
[3] Singh AK and Gupta B 2015 A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram. *Procedia Computer Science* 54 676-682.
[4] Huang Q 2020 Segmentation of breast ultrasound image with semantic classification of superpixels. *Medical Image Analysis* 61 101657.
[5] Lakhani SR, Ellis IO, Schnitt SJ, Tan PH and Vijver MJ van de (Eds.) 2012 *WHO Classification of Tumours of the Breast* (4th ed.). International Agency for Research on Cancer (IARC).
[6] Qu AP, Chen JM, Wang LW, Yuan JP, Yang F, Xiang QM, Maskey N, Yang GF, Liu J and Li Y 2015 Segmentation of Hematoxylin-Eosin stained breast cancer histopathological images based on pixel-wise SVM classifier. *Science China Information Sciences* 58 1–13.
[7] Fouad S, Randell D, Galton A, Mehanna H and Landini G 2017 Unsupervised Superpixel-
Based Segmentation of Histopathological Images with Consensus Clustering. *Medical Image Understanding and Analysis* **723** 767–779.

[8] Xu J, Luo X, Wang G, Gilmore H and Madabhushi A 2016 A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* **191** 214–223.

[9] Pan P, Chen H, Li Y, Cai N, Cheng L and Wang S 2020 Tumor Segmentation in Automated Whole Breast Ultrasound Using Bidirectional LSTM Neural Network and Attention Mechanism. *Ultrasonics* 106271.

[10] Punitha S 2018 Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal* **3** 348-358, 2018.

[11] Aswathy MA 2020 Performance Analysis of Segmentation Algorithms for the Detection of Breast Cancer. *Procedia Computer Science* **167** 666-676.

[12] Rodtook A 2013 Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer. *Journal of Visual Communication and Image Representation* **24** 1414-1430.

[13] Macqueen J 1967 Some Methods for Classification and Analysis of Multivariate Observations. *Fifth Berkeley Symp. on Math. Statist. and Prob* **1** 281–297.

[14] Al-shammaa AAM and Mohamed HR 2012 Extraction of connected components Skin pemphigus diseases image edge detection by Morphological operations. *International Journal of Computer Applications* **46** 7-13.

[15] Yee Jing T, Mustafa N, Rahman KSA 2021 Segmentation of Tumour Regions for Tubule Formation Assessment on Breast Cancer Histopathology Images. *International Conference on Robotics, Vision, Signal Processing and Power Applications.*