Standard Deviation Based Otsu for Breast Cancer Classification

Siti Salmah Yasiran[1]*, Shaharuddin Salleh[1], Rozi Mahmud[2]

[1]*Center for Industrial and Applied Mathematics, Faculty of Sciences,*
Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia
[2]*Cancer Research and Education Centre, Faculty of Medicine and Health Sciences,*
Universiti Putra Malaysia, Serdang, Selangor, Malaysia

Segmentation is one of the essential components in Computer Aided Diagnosis (CADx). This paper proposed a modified Otsu method for segmentation phase in classifying breast cancer. In Otsu's method, the process of computation within-class variance for two classes produce lower accuracy for the classification, hence the standard deviation will be utilized to overcome this limitation. After that, the features of the segmented images will be extracted by using Speed Up Robust Features (SURF). Finally, the corresponding features will be classified using several classifiers. It is found that the Support Vector Machine (SVM) classifier shows the highest accuracy rate among other classifiers. Moreover, experimental results show that the proposed method is superior to the original Otsu method.

Keywords: Otsu, Standard Deviation, Mathematical Morphology, breast cancer, Computer Aided Diagnosis (CADx), Support Vector Machine (SVM).

I.   INTRODUCTION

Computer Aided Diagnosis (CAD) has been widely used to assist physicians in analysing their findings especially in breast cancer. According to Manoharan *et. al.,* (2010), the mammogram is considered as the standard method to detect the abnormalities of the breast at a primary stage. There are two stages involved in the whole mammogram procedure. The first stage is screening procedure stage and the last stage is diagnostic mammography. The role of CAD occurs in the diagnostic mammography stage. It has been proven that the CAD able to reduce the misevaluated findings by physicians as well as to increase the performance of the breast cancer classification (Li & Zhou, 2007). The Computer Aided Diagnosis component (CADx) is on the key component inside the CAD. The role of CADx is essentials in finalizing the type of treatment that should be taken by a patient. There are five phases included in the CADx component.

The first phase is the pre-processing, this is followed by the segmentation phase. The third and fourth phase is features extraction and features selection respectively. The final phase is the classification phase. In this paper, the main contribution is focus on the segmentation phase of the CADx component. This is an extended works from Salleh et al., (2017), whereby the segmentation phase is not included in their work.

Image segmentation is the most crucial steps in the analysis of image, especially to determine the performance of the CAD. The purpose of image segmentation is to split an image into regions that have a strong connection within the image objects. The pixel intensity refers to the simplest property that can be found in an image. Thus, the segmentation method through the intensity thresholding is selected in the segmentation phase (Somasundaram & Genish, 2012). One of the benefits of the thresholding is, it

*Corresponding author's e-mail: sitisalmah@tmsk.uitm.edu.my

helps to create binary images from grey level image by twisting all pixels below a threshold to zero and the remaining pixels to one (Sonka *et. al.,* 1993). Various methods have been proposed by researchers to binarize an image (Tabbone & Wendling, 2003; Sauvola & Pietikäinen, 2000). In this paper, we propose a modified Otsu`s method to binarize an image. This is because the binarization process will reduce the computation costs. It is known that the process of computation within-class variance for two classes is high.

II.   MATERIALS AND METHODS
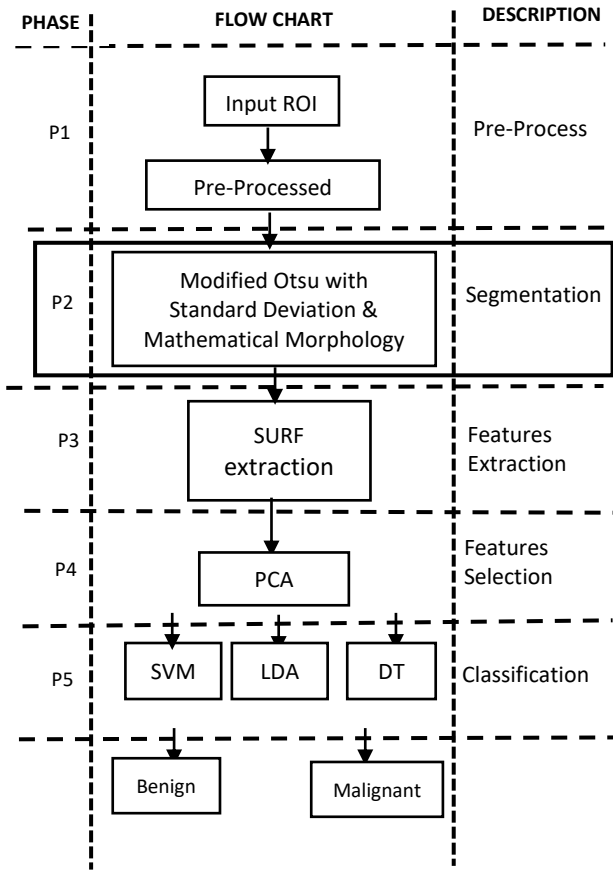
The proposed method is represented as in Figure 1.

Figure 1: Breast cancer classification.

From Figure 1the proposed method is divided into five phases. The first phase is pre-process, followed by the segmentation phase. Phase 3 and phase 4 is the features extraction and features selection respectively. Finally, the classification phase is to classify either the cancer is benign or malignant. The description of Phase 1, Phase 3, Phase 4 and Phase 5 is already explained thoroughly by Salleh *et. al.,* (2017). This paper is focusing on the segmentation part. The details explanation of phase 2 is described in the next section.

A.   Modified Otsu Thresholding

Otsu's method is composed of iteration process through all possible threshold values and compute a measure of spread for the pixel levels each side of the threshold. The target is to find the threshold value where the total of foreground and background is minimum. Let ω be the weighted total of the variances of object and β is the background in the within-class variances. It can be represented mathematically as;

$$\sigma_W^2(I) = \sum_{i=0}^{I-1} \rho(i) \cdot \sigma_\beta^2(I) + \sum_{i=I}^{N-1} \rho(i) + \sigma_W^2(I] \quad (1)$$

where $\sigma_\beta^2(I)$ refers to the variance of the pixels in the background of image I . The term $\sigma_W^2(I)$ is the variance of the pixels in the foreground. The term $\rho(i)$ is the probability of pixel value occurrence. From Equation (1), the between-class variance is obtained by;

$$\sigma_\psi^2(I) = \sigma^2(I) - \sigma_\varsigma^2(I) \quad (2)$$

where $\sigma_\psi^2(I)$ is the between-class variance of image I. The term $\sigma^2(I)$ refers to the combined variance and mean of the pixels and replacing the mean $,m$ into Equation (1) yields;

$$\sigma_\psi^2(I) = \sum_{i=0}^{I-1} \rho(i) \cdot \left[m_\beta(I) - m\right]^2 +$$
$$\sum_{i=I}^{N-1} \rho(i) \cdot [m_W(I) - m]^2 \quad (3)$$

where $m_\beta(I)$ is the mean of the pixels background of image I and the term $m_W(I)$ is the mean of the pixels of the object of image I.

To solve Equation (3), it requires high computational costs. In the Original Otsu method, the mean value is utilized to solve Equation (3). However, it produces lower accuracy of the classification. Thus, the standard deviation is utilized instead of the mean value.

The equation of the standard deviation is ;

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}} \quad (4)$$

where the term N is the pixels of intensity x_i with $i = 1,2, \dots N$. The terms $\bar{x}$ is the mean of x_i. Then, by replacing Equation (4) into Equation (3), we will obtain;

$$\sigma_\psi^2(I) = \sum_{i=0}^{I-1} \rho(i) \cdot \sum_{i=I}^{N-1} \rho(i) \cdot [SD_\beta - SD_\omega]^2 \quad (5)$$

where the term SD_β represent the standard deviation of the background image I. The term SD_ω is the standard deviation of the object image I. The optimum threshold value will be computed by using Equation (5). Figure 2 illustrates the algorithm of our proposed method.

Algorithm: Modified Otsu
1:        Calculate Histogram and probalities, H_{max}
2:        **let $T_\circ$ be the increment potential threshold $\ni$** $I = T_\circ, T = I, \&\&S_{max} = 0;$
3:        **while $I < H_{max}$ do**
4:           Calculate $\sum_{i=0}^{I-1}\rho(i)\&\&\sum_{i=I}^{N-1}\rho(i)$
5:           Calculate $SD_\beta\&\&SD_\omega$
6:           **if** $\sigma_\beta^2(I) < S_{max}$
7:           **then $S_{max} = \sigma_\beta^2(I)\&\&, T = I;$**
8:           **end if**
9:        **set $I = I + T_\circ,$**
10.   **end while**

Figure 2: The Algorithm of modified Otsu.

From figure 2, the algorithm explains the process of modified Otsu method, starting from the Equation (1) until Equation (5).

In this paper, the rule of 70% of the data used for training and 30% of the data used for testing is utilized (Forman & Cohen, 2004). After the pre-process phase, the threshold value of each image is obtained by using the modified Otsu through Equation (5). Then the corresponding threshold value will be utilized to segment the corresponding area of benign and malignant images by using the mathematical morphology.

B. *Mathematical Morphology*

In image analysis, the mathematical morphology method utilizes the concept of set theory (Kimori, 2011). The outline and the structure of the image will be analysed through these methods. In mathematical morphology, the role of structuring elements (SE) is to transform the original image into certain shape and size through various iterations. There are two simple operations involved in mathematical morphological; erosion and dilation. Mathematically, the erosion operator can be represented as;

$$A \ominus B = \{Z|(B)_Z \subseteq A\}, \ A, B \in Z^2 \qquad (6)$$

where A refers to test image and B is the corresponding SE. Both A and B are in the set of Z^2. Equation (6) stated that the erosion of A by B translated is contained in A. Meanwhile, the dilation operators can be expressed as;

$$A \oplus B = \{Z|(B)_Z \cap A \neq \emptyset\}, \ A, B \in Z^2 \qquad (7)$$

where the term $\emptyset$ refers to a null set. Equation (7) is based on reflecting B in origin and shifting this reflection by Z. The dilation of image A by SE of B is the set of all displacements. This paper utilized the erosion operators for the segmentation.

III. RESULTS AND DISCUSSION

The same benchmark for performance measurement used by Salleh *et. al.,* (2017) will be utilized in this paper. The results of both original Otsu and the modified Otsu is tabulated as in Table 1.

Table 1. Results of original and modified Otsu

	Classifiers	Accuracy	Az	FPR(%)	FNR(%)
Original Otsu	SVM	81.1	0.88	10.5	0.13
	LDA	45.3	0.48	1.01	0.97
	DT	41.0	0.45	1.1	0.95
Modified Otsu	SVM	85.9	0.91	11.7	0.10
	LDA	51.3	0.51	1.15	0.9
	DT	53.1	0.53	1.79	0.88

From Table 1, the table is composed into two main rows. The first main row represents the original Otsu method, and the second main row represents the modified Otsu method. It can be observed that the highlighted values for both main rows indicate that the SVM classifier is superior compared to the other classifiers. The accuracy of the modified Otsu method is increasing by 4.8% compared to the original Otsu method. Figure 3 illustrates the Area Under the Curve (AUC) for all classifiers.
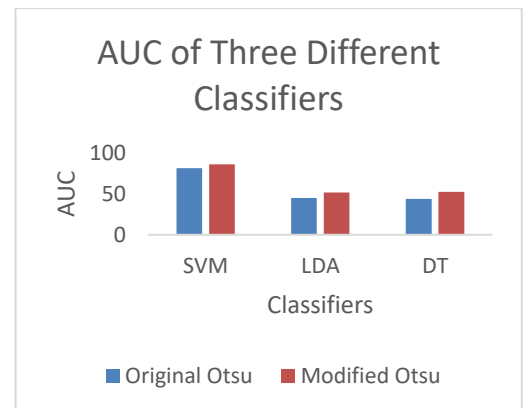

Figure 3. The AUC of all classifiers.

From Figure 3, there are three classifiers used in this paper; Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Decision Tree (DT). It can be observed that all graph shows an increasing in accuracy for the modified Otsu method. For the LDA classifier, there are 6.0% increment from the original Otsu method. Meanwhile, there are 2.1% increment for the DT classifier. In this paper, the Fine Gaussian kernel of SVM is utilized for classification phase. This utilization is based on findings from Salleh *et. al.,* (2017). It can be concluded that there is progressive increment in the performance of the CAD through the modification of the original Otsu method.

IV. CONCLUSION

We have proposed a modified Otsu method to classify the breast cancer, either it is benign or malignant. The main contribution of this paper occurs at the segmentation phase. Instead of mean value formula from the original Otsu method, the standard deviation is utilized to find the optimum threshold value of each corresponding image. Then, the mathematical morphology method is implemented to segment the corresponding boundaries of benign and malignant. The erosion operator is used to erode the boundaries of foreground pixels.

Moreover, the process of the whole CAD is fully automatic. It is recommended that different type of statistical method is utilize for the modification of Otsu method. This is important to increase the performance of the whole CAD system as well as to reduce the error rate did by physicians. Thus, the mortality rate among the breast cancer patients can be reduced.

V. REFERENCES

C. Manoharan, N. V. S. Sree, and R. Lakshmi, 2010, "Classification of Micro Classifications in Mammogram using Combined Feature Set with SVM,".

G. Forman and I. Cohen, 2004, "Learning from Little: Comparison of Classifiers Given Little Training," Springer, Berlin, Heidelberg,pp 161-172.

J. Sauvola and M. Pietikäinen, 2000, Feb "Adaptive document image binarization," *Pattern Recognit.* 33, (2), pp. 225–236.

K. Somasundaram and T. Genish, 2012, "Modified Otsu Thresholding Technique," Springer, Berlin, Heidelberg, pp. 445–448.

M. Li and Z.-H. Zhou, 2007, "Improve Computer-Aided Diagnosis with Machine Learning Techniques Using Undiagnosed Samples," *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans.* 37, (6), pp. 1088–1098, Nov.

M. Sonka, V. Hlavac, and R. Boyle, 2012, "Image Processing, Analysis, and Machine Vision Second Edition,"Springer, Berlin, Heidelberg.

S. Tabbone and L. Wendling, 2003, "Multi-scale binarization of images," *Pattern Recognit. Lett.*24, (1–3), pp. 403–411, Jan.

S. Salleh, R. Mahmud, H. Rahman, S. Y.-J. of F. and, and undefined 2017, "Speed Up Robust Features (Surf) With Principal Component Analysis-Support Vector Machine (Pca-Svm) For Benign And Malignant," *jfas.info.*

Y. Kimori, 2011, "Mathematical morphology-based approach to the enhancement of morphological features in medical images.," *J. Clin. Bioinforma.*1, p. 33, Dec.