Otsu and Mathematical Morphology for Breast Cancer Classification

Siti Salmah Yasiran[1*], Shaharuddin Salleh[1] and 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 combination of Otsu and mathematical morphology for segmentation phase in classifying breast cancer. Then, 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 combination method shows promising results.

Keywords: Otsu, Mathematical Morphological, Speed Up Robust Features (SURF), 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 Mahonaran 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 (Somasundram & Genish, 2012). One of the benefits of the thresholding is, it 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.,* 2014).

In this paper, we propose a combination of Otsu method and the mathematical morphology method for the segmentation phase of the CADx. The mathematical morphology is utilized since it simplifies image data, preserve their essential shape characteristic and eliminate noise.

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

II.   MATERIALS AND METHODS

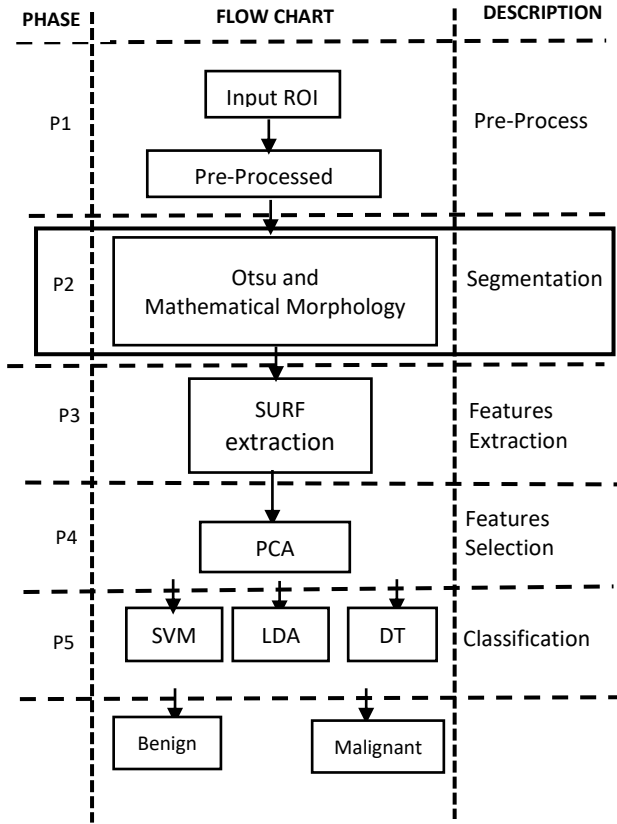The proposed method is illustrated in Figure 1.

Figure 1. The proposed method

From Figure 1, there are three main columns that represent the phase of the proposed method, this is followed by the second column, the flow chart of the proposed method and the final column which represent the description of the flow chart for each phase. The proposed method is divided into five phases. 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.   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 (Somasundram & Genish, 2012)

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.

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 is completed, the threshold value of each image is obtained by Otsu method. 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;

$$A \ominus B = \{Z | (B)_Z \subseteq A\}, \ A, B \in Z^2 \quad (4)$$

where A refers to test image and B is the corresponding SE. Both A and B are in the set of Z^2. Equation (4) stated that the erosion of A by B translated is contained in A. Meanwhile, the dilation operators can be expressed as;

$$A \oplus B = \{Z | (B)_Z \cap A \neq \emptyset\}, \ A, B \in Z^2 \quad (5)$$

where the term Ørefers to a null set. Equation (5) 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 DISCUSSIONS

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

Table 1. Results of the Otsu and combination Otsu with mathematical morphology

	Classifiers	Accuracy	Time (sec)
Otsu only	**SVM**	**80.6**	**7.49**
	LDA	55.69	11.5
	DT	48.7	8.64
Otsu with MM	**SVM**	**84.9**	**6.53**
	LDA	57.86	10.4
	DT	57.3	7.11

From Table 1, the table is composed into two main rows. The first main row represents the Otsu method only and the second main row represents the combination of Otsu method with mathematical morphology. 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 combination Otsu method with mathematical morphology is increasing by4.3% compared to Otsu method only. the Figure 3 illustrates the Area Under the Curve (AUC)for all classifiers.

From Figure 2, 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 combination Otsu method with mathematical morphology. For the LDA classifier, there are2.17% increment. Meanwhile, there are 8.6% increment for the DT classifier. The Fine Gaussian kernel of SVM is utilized for classification phase. This utilization is based on findings from Salleh *et. al.,* (2017). Figure 3 represents the malignant image results

of phase 1 until phase 3 of the proposed method of the whole CAD system.
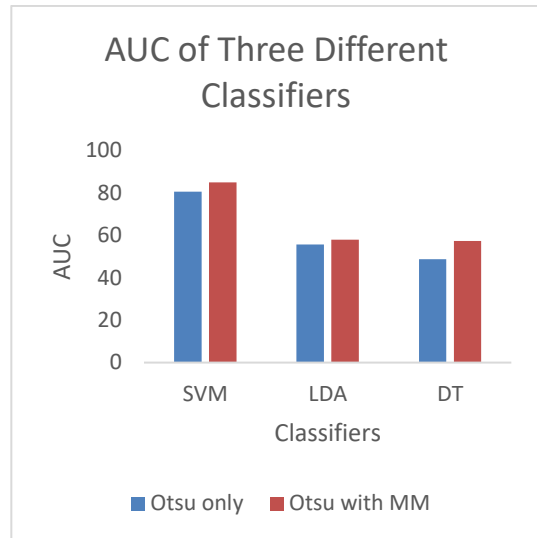

Figure 2. The AUC of all classifiers

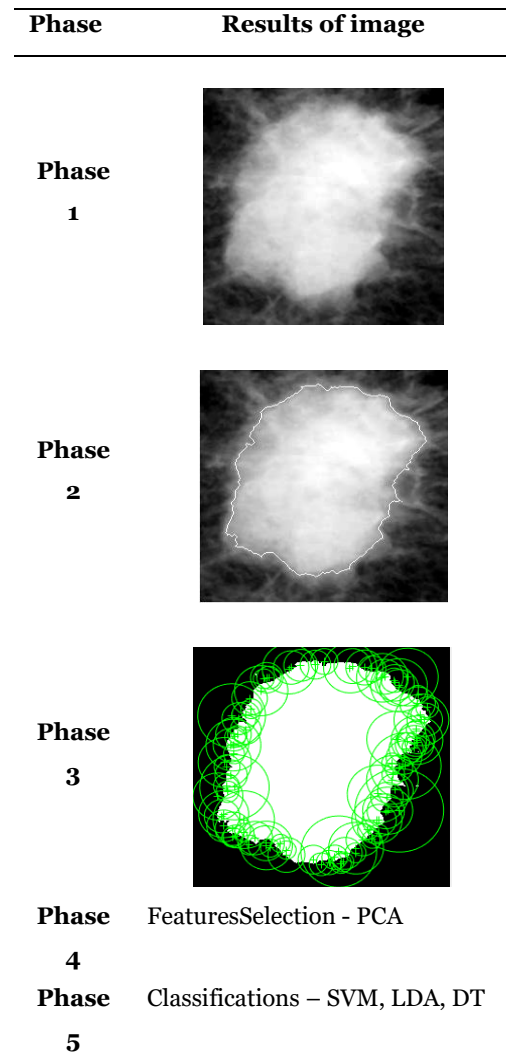| Phase	Results of image
Phase 1	
Phase 2	
Phase 3	
Phase 4	FeaturesSelection - PCA
Phase 5	Classifications – SVM, LDA, DT

Figure 3. The results of malignant Image

From Figure 3, the phase 1 shows the pre-processed image of the malignant image. This is followed by the result of phase 2, our proposed method for image segmentation. The white line represents results of the combination Otsu and mathematical morphology. The segmentation illustrates the boundaries of the corresponding malignant image. Finally, the phase 3 shows the results of the SURF descriptors. It is based on the results of the segmented image in phase 2. The circles indicate the strongest point of the SURF descriptors. Then, the features values from the SURF descriptors will be extracted and stored in an array for the next phase. After that, the features selection phase will be carried out to reduce the dimension of redundancies and to increase the accuracy of the CADx. Finally, the classification phase will be conducted to classify either the cancer is benign or malignant. There are three different classifiers used in this paper: Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Decision Tree (DT) for comparison purpose.

IV. CONCLUSION

We have proposed combination of Otsu thresholding and mathematical morphology method in the segmentation phase of the CAD. The mathematical morphology method is implemented to segment the corresponding boundaries of benign and malignant. Moreover, the process of the whole CAD is fully automatic. It is recommended that the modification is focus on the Otsu method itself to increase the performance of the whole CAD system.

V. REFERENCES

Forman, G. & Cohen, I. 2004, Learning from Little: Comparison of Classifiers Given Little Training. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 3202, pp.161-172

Kimori, Y. 2011, Mathematical Morphology-Based Approach To The Enhancement Of Morphological Features In Medical Images. Journal of Clinical Bioinformatics.1 (33), pp. 1-10.

Li, M.&Zhou, Z.H. 2007, Improve Computer-Aided Diagnosis With Machine Learning Techniques Using Undiagnosed Samples. IEEE Transaction System. 37(6), pp. 1088–1098.

Manoharan, C., Sree, N. V. S.&Lakshmi. R. 2010, Classification of Micro Calcifications in Mammogram using Combined Feature Set with SVM. International Journal of Computer and Application. 11(10), pp. 30-34.

Salleh, S., Mahmud, R., Rahman, H.& Yasiran, S. S. 2017, Speed Up Robust Features (SURF) With Principal Component Analysis-Support Vector Machine (PCA-SVM) For Benign and Malignant Classifications. Journal of Fundamental and Applied Science.9(5), pp. 624-643.

Somasundaram, K. & Genish, T. 2012, Modified Otsu Thresholding Technique. International Conference on Mathematical Modelling and Scientific Computation. 283, pp. 445–448.

Sonka, M., Hlavac,V.& Boyle, R. 2007, Book Chapter. Image Processing, Analysis, and Machine Vision Second Edition.