

**SPEED UP ROBUST FEATURES (SURF) WITH PRINCIPAL COMPONENT ANALYSIS-SUPPORT VECTOR MACHINE (PCA-SVM) FOR BENIGN AND MALIGNANT CLASSIFICATIONS**

S. Salleh<sup>1</sup>, R. Mahmud<sup>2</sup>, H. Rahman<sup>3</sup> and S.S. Yasiran<sup>1,\*</sup>

<sup>1</sup>Center for Industrial and Applied Mathematics, Faculty of Sciences, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

<sup>2</sup>Cancer Research and Education Centre, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

<sup>3</sup>Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Published online: 17 October 2017

---

**ABSTRACT**

A novel Computer Aided Diagnosis (CADx) component is proposed for breast cancer classifications. Four major phases were conducted in this research. The first phase is pre-processing, this is followed by features extraction phase by using the Speed Up Robust Features (SURF). The next phase is features selection by using the Principal Component Analysis (PCA). The final phase is the classification phase to classify the cancer. Three different classifiers; Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Decision Tree (DT) were compared in this research. Results obtained shows that the PCA-SVM performs the highest accuracy with 92.9% accurate compared to other classifiers.

**Keywords:** breast cancer; CADx; SURF; PCA; SVM.

---

Author Correspondence, e-mail: [sitisalmah@tmsk.uitm.edu.my](mailto:sitisalmah@tmsk.uitm.edu.my)

doi: <http://dx.doi.org/10.4314/jfas.v9i5s.44>



## 1. INTRODUCTION

Currently, an approximation of 1.66 million new cancer cases was projected to occur in the United State in 2015. The projection is made for all types of cancer. It is known that the breast cancer is the second leading cause of death around the world. Based on this [1]projection, it was reported that nearly 230,000 new breast cancer cases in the year 2013 at the same country [2]. Hence, an early detection of breast cancer need to be conducted in order to reduce the number of new breast cases as well as the mortality cases of breast cancer. According to [3], the reliable method to detect the breast cancer at initial stage is by using mammogram. The whole process of mammogram procedure is divided into two phase; the screening procedure and the diagnostic mammography.

In normal practice, a total of four mammogram images of a breast cancer patient will be displayed after the screening procedure is completed. These images will include the Cranial Caudal (CC) view for both left and right view and the Mediolateral Oblique (MLO) view, also for both left and right view for each patient. Thus, these procedure will produce numerous numbers of mammogram needed to be evaluated by an expert radiologist. The evaluation will be done thru visual eye bowling. By considering the human factor, the detection of potential cancerous might be misevaluated (error). In this situation, it is quite challenging for the radiologist to perform accurate evaluation to detect the potential cancer locations. According to [4], there will be 10% to 30% error occur during the detection process.

After the screening procedure is complete, the diagnostic mammography stage will be carried out by the radiologist. In this stage, the radiologist will diagnose either the cancer is classified as a benign or malignant. Similarly, human error also occurs during this stage. There are numerous of studied have been done to overcome this error. The Computer Aided Diagnosis (CAD) has been proved to reduce the error rate as well as to increase the performance of the breast cancer detection and classification [5].

The Computer Aided Diagnosis (CADx) component is one of the main component in the CAD. In general, the CADx component occur in the diagnostic mammography stage to assist radiologist in making devaluation either the cancer is benign or malignant. If the cancer is classified as benign then the radiologist will decide either the biopsy procedure is essential to the corresponding patient or not. The role of the CADx component is crucial in determine the

type of treatment to the patient as well as to reduce the mortality rate. The CADx component is composed of four sub phases; pre-processing, features extraction, features selection and classification. In this paper, a novel CADx component will be proposed to assist radiologist in the diagnostic mammography stage either the cancer is classified as a benign or malignant.

This paper is organized as follows. Section 1.1 will briefly explain the related works of our proposed CADx component. Then, section 2 will details out the method in the CADx development. The related mathematical formulation will be explained in each of the sub phases in the CADx component. The results will be discussed at section 3 and the final section conclude all of our findings.

### **1.1.Related Works**

The CAD for mammography is firstly introduced at the year of 1990's [6]. A study conducted by [7] was successfully developed a new group of features to enhance the performance of CADx system. The Radial Distance vector of masses were extracted and transformed to the frequency domain to produce new features called the Fourier Transform Radial Distance (FTRD). The performance of the classifiers with proposed features were measured based on the ROC curve with accuracy of 98%. Another study which focus on the method of moment was conducted by [8]. Two method of moments: Zernike and Krawtchouk were used and compared for the mass characterization. It is found that the Krawtchouk moment produce the best accuracy rate compared to the Zernike moment with 90.2% and 81% respectively.

A novel approach for automatic microcalcification and cluster detection CAD system was developed by [9]. The morphology features of microcalcification were used as the feature extraction and the boost classifiers were employed in their study. They have suggested to conduct further study with bigger database is crucial to prove the capability of the proposed method. Along the same line, in [10] have developed automatic CAD system to classify four different tissue density. Two-dimensional principal component analysis (2DPCA) were used to extract tiny patches of ROI. They have concluded that the CBIR method reach accurate rates to assist radiologists in classifying breast density.

In [11] have introduced a novel CAD system to detect breast masses based on fusion scheme of features. They have proposed a new scheme of SVM fusion based on diversity and complementary between features. Diversity criteria between features are adopted in their study

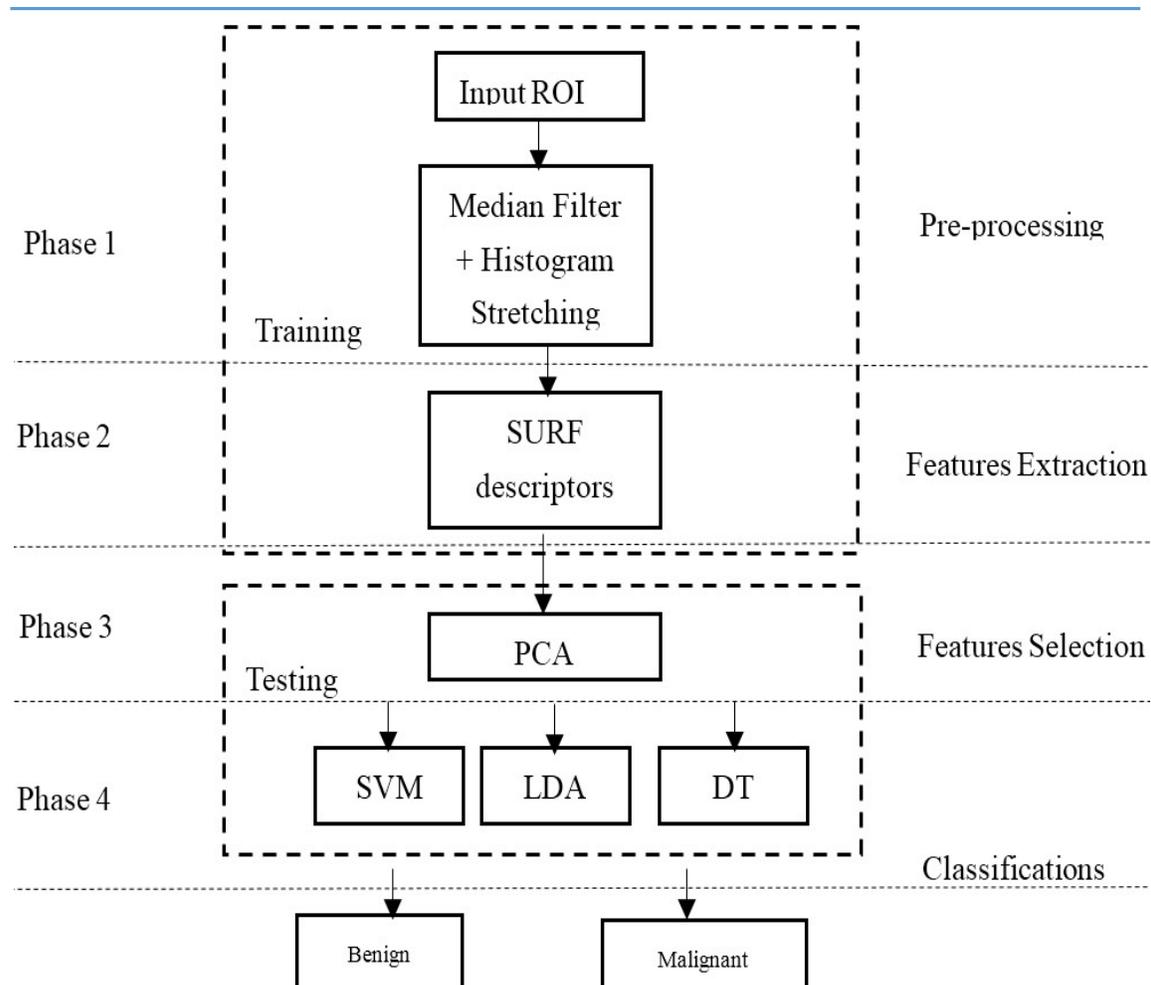
to ensure best performance. The results shows that the proposed system are accurate. A mammogram classification scheme have been proposed by [12]. The Gray level Co-Occurrence Matrix was utilized for the features. The MIAS and DDSM database were implemented in their study. It is found that the proposed scheme give promising results for both database.

In [13] have proposed two automatic methods for masses diagnosis. The Cellular Neural Network (CNN) was implemented as the segmentation method with the utilization of the Genetic Algorithm (GA) for the parameter determination. The proposed method produces high rates in sensitivity, specificity and also accuracy. In [14] have successfully improved the Mann-Whitney statistical test for features selection in reducing the dimensionality of features in the CADx. They found that the redundancy analysis is a crucial phase to increase the performance of the classification.

Recently, in[15] have proposed a CADx based on the Haralick and Invariant Moment Features. The SVM was utilized as the classifiers to classify two categories. The first category is either the cancer is benign or malignant. The other category is to classify the character of background breast tissue. The accuracy of the proposed features is 90.5% and 77.5% respectively.

## **2. MATERIALS AND METHODS**

In this research, the proposed CADx component is presents as in Fig. 1.



**Fig.1.** The proposed CADx component

From Fig. 1, the CADx component is composed of four main phases; pre-processing, features extraction, features selection and finally the classification phase. There are two boxes which indicate where training and the testing phase occurs. The explanation of each phases of the proposed CADx is discussed in the following subsection.

### 2.1. Pre-Processing

In this phase, the data is taken from the MIAS database. The database image has been reduced to a 200 micron pixel edge with fixed size; 1024 by 1024 pixels. The images is composed of normal breast tissue, benign and malignant. The benign is composed of six abnormalities which are calcifications, circumscribed masses, other ill-defined masses, architectural distortion, anasymmetry and speculated masses. The ground truth location of the cancer will be confirm by radiologist for the Region of Interest (ROI) purpose. Then, the ROI will be resize to a fixed standard size.

Next, the image enhancement technique will be applied to remove the noise and to enhance the ROI image. In this research, the combination of median Filter and Histogram stretching technique [16-17] will be applied in each ROI image. The median filter will be selected since it considers each pixel in the image and find its neighbors to conclude either it represent its surrounding or not. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values.

This can be expressed as

$$q(x, y) = \underset{(v,w) \in S_{xy}}{\text{Med}}(r(v, w)) \quad (1)$$

where  $q$  is the function of a point  $(x, y)$ . The term  $S_{xy}$  is the set of coordinates in the rectangular sub image centred at a point  $(x, y)$ . The median function calculates the median of the image  $r(v, w)$  in the area defined by  $S_{xy}$ . The Histogram Stretching is also known as contrast stretching. The pixels with grey levels greater than 255 will be substituted with 255. Conversely, any pixel with grey level less than the minimum grey level value will be replaced with 0. The Histogram Stretching equation can be expressed as

$$I(x, y) = 255 \cdot \frac{I(x, y) - \text{Min}}{\text{Max} - \text{Min}} \quad (2)$$

where  $I(x, y)$  is the new grey level of the pixel image, meanwhile Max and Min represent the maximum and minimum grey level of the image respectively. The operation is applied in two phases. In the first phase, the minimum and maximum intensity values of the image are calculated. Then, the above formula is applied on each pixel.

## 2.2. Features Extractions

Speed Up Robust Features (SURF) was proposed by [18], to overcome the weaknesses of the Scale Invariant Feature Transform (SIFT). The SURF method is utilized in this paper due to its robust behaviour[19] to scaling, blurring, orientation, illumination and also viewpoint changes. In addition to that, computational speed can also be improved through the application of SURF method. SURF have been applied in various field such as image processing and video tracking. [20-22].

In general, the algorithm of SURF is presents as in Fig. 2.

**Algorithm.** Speed Up Robust Features (SURF)

---

```

1: Input:  $p(x, y)$ 
2: for each image,  $\forall(x, y)$  do
3:   Convert to  $p$  grayscale,  $I$ ;
4:   Create Integral Image,  $S$ ; //create
   integral image
5: end for
6: for each  $S \in L$  do
7:   Compute Hessian Matrix,  $H^L$  //
   Hessian Detector
8:   Convolve  $H^L$  with Gaussian
   function,  $G^*(\sigma)$ 
9:   Find  $\Gamma_{xx}^L, \Gamma_{yy}^L, \Gamma_{xy}^L$  //second order
   box filter
10:   if  $(\Gamma_{xx}^L \& \& \Gamma_{yy}^L) \geq 0 \& \& (\Gamma_{xy}^L < 0)$ 
11:     Local maxima and  $P_{xx}, P_{yy}, P_{xy}$ 
   obtained
12:   end if
13:   Find  $|H_{approx}|$  //detect features points
14: end for
15: for  $\forall(x, y) \rightarrow x, y$  directions do
16:   Find Haar wavelet,  $d_x$  and  $d_y$  //generate descriptors
17: end for
18: Output: SURF features

```

---

**Fig.2.**The SURF algorithm

Based on Fig. 2, the first steps in the SURF algorithm is the integral image creation. Integral image can be expressed by the integral image of a point  $I(x, y)$  where  $I$  is the grayscale image and  $(x, y)$  is the coordinate pixel. The point can be used to represent the sum of rectangle origin in the original image is mathematically described as follows

$$S(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (3)$$

where  $S(x, y)$  is the the total integrated original image. The terms  $I(i, j)$  is the pixel value of the original image. The next step is the Hessian Detector Filters due to its robustness to reduce the computational time. It is also responsible to identify extreme points on each layers on the image scale space through the Hessian Matrix [23] as follows

$$H^L(x, y, \sigma) = \begin{bmatrix} \Gamma_{xx}^L(x, y, \sigma) & \Gamma_{xy}^L(x, y, \sigma) \\ \Gamma_{xy}^L(x, y, \sigma) & \Gamma_{yy}^L(x, y, \sigma) \end{bmatrix} \quad (4)$$

where  $H^L$  is the Hessian Matrix at scale space  $L$ . Then, by convolving the image point with the two-dimensional Gaussian function,  $G^*(\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ , we will obtain the following equations

$$\begin{aligned} \Gamma_{xx}^L(x, y, \sigma) &= \frac{\partial^2 G^*(x, y, \sigma)}{\partial x^2} \\ \Gamma_{yy}^L(x, y, \sigma) &= \frac{\partial^2 G^*(x, y, \sigma)}{\partial y^2} \\ \Gamma_{xy}^L(x, y, \sigma) &= \frac{\partial^2 G^*(x, y, \sigma)}{\partial x \partial y} \end{aligned} \quad (5)$$

where  $\Gamma_{xx}^L(x, y, \sigma)$ ,  $\Gamma_{yy}^L(x, y, \sigma)$  and  $\Gamma_{xy}^L(x, y, \sigma)$  are the second order box filters at scale space  $L$ , which refer to the results after the convolution process results. This is also known as Laplacian of Gaussians. The local maxima of this filter appear in regions if both  $\Gamma_{xx}^L$  and  $\Gamma_{yy}^L$  are positive and  $\Gamma_{xy}^L$  negative. It can be represented by images with the highest intensity gradient variations in various directions. The parameter  $\sigma$  in the  $G^*(\sigma)$  is known as sigma. Then, the detection of features point is carried out. In this step, the results of the previous step produce the determination of approximate Hessian Matrix which can be expressed as;

$$|H_{approx}| = P_{xx} \cdot P_{yy} - (\kappa P_{xy})^2 \quad (6)$$

where  $|H_{approx}|$  is the determinant of the approximation value of the Hessian Matrix. This determinant calculates all positions of the integral images. The terms  $P_{xx}$ ,  $P_{yy}$  and  $P_{xy}$  are the results of the convolution with  $\Gamma_{xx}^L$ ,  $\Gamma_{yy}^L$  and  $\Gamma_{xy}^L$  respectively. The term  $\kappa$  refers to the weighting parameter, representing the energy conservation between the  $G^*(\sigma)$  and the approximate  $G^*(\sigma)$ . To detect the feature points, the pyramid scale space will be constructed. It will directly change the scale of the box filters to utilize the scale space [24]. The final step is the descriptors generation. The orientation information of the feature point is also generated in

this step. The features point is generated by counting the Haar wavelet in  $x$  and  $y$  direction, which can be denoted as  $d_x$  and  $d_y$ . Hence, all values in the subregion is expressed as

$$\sum d_x \text{ and } \sum d_y \quad (7)$$

and by calculating the sum of absolute value in both Equation (7), it represent the responsive value changing in the  $x$  and  $y$  direction of the Haar wavelet. Next, the features selection phase is carried out.

### 2.3. Features Selection

The Principal Component Analysis (PCA) is utilized due to its capability to reduce the dimensionality of the input features [25], as well as to increase the performance of the proposed CADx. In addition to that, the PCA also capable to avoid redundancy in features extraction phase. The algorithm of the PCA is illustrates as in Fig. 3.

---

#### Algorithm. Principal Performance Analysis (PCA)

---

<p>1: <b>Input:</b> <math>v = \{a_1, a_2, a_3 \dots a_v\}</math> // Training dataset</p> <p>2: <math>n</math> = dimension space</p> <p>3: <b>for</b> each vectors, <math>v_i \in n \quad \forall v</math>, <b>do</b></p> <p>4: Find the mean, <math>\bar{v}</math>;</p> <p>5: Subtract <math>\bar{v}</math> for each <math>v, \left\{v - \bar{v}\right\}_n</math></p> <p style="padding-left: 20px;">//Standardize dataset</p> <p>6: <b>end for</b></p> <p>7: <b>for</b> each vectors, <math>\bar{v}_i \in n \quad \forall \bar{v}</math>, <b>do</b></p> <p style="padding-left: 20px;">Find the covariance, <math>CV</math> ; // Create covariance matrix</p>	<p>9: Find eigenvalues, <math>\mu_1, \mu_2, \dots \mu_n</math>;</p> <p>10: Find eigenvectors, <math>\lambda_1, \lambda_2, \dots \lambda_n</math>;</p> <p>11: <b>end for</b></p> <p>12: <b>while</b> (<math>\lambda_n \geq \lambda_{n+1}</math>) &amp;&amp; (<math>\tilde{n} \leq n</math>) // Find largest <math>\lambda</math></p> <p>13: Transform <math>v</math> to <math>\tilde{v}</math>, such that ;</p> <p>14: <math>\{a_1, a_2, a_3 \dots a_v\} \rightarrow \{\tilde{a}_1, \tilde{a}_2, \tilde{a}_3 \dots, \tilde{a}_v\}</math> &amp; <math>n \rightarrow \tilde{n}</math></p> <p>15: <b>end while</b></p> <p>16: <b>Output:</b> Transformed dataset <math>c_{l,i}</math></p>
--	---

---

**Fig.3.**The PCA algorithm

From Fig. 3, there are four basic steps in the PCA algorithm. The first step in PCA is to standardize the  $v$  vectors dataset  $(a_1, a_2, a_3 \dots a_v)$  from  $n$  dimensional space. These step can be expressed as

$$\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i \quad (8)$$

where  $\bar{v}$  is the mean of the vectors  $v$ . Then, the mean of each dimension will be subtracted.

In the second step, the covariance matrix is calculated by using

$$CV = \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})(v_i - \bar{v})^T \quad (9)$$

where  $CV$  is the covariance matrix and  $T$  is the matrix transpose. In the next step, Equation (9) produces eigenvectors eigenvalues and PCA transform vectors  $v$  to  $\tilde{v}$  vectors  $(\tilde{a}_1, \tilde{a}_2, \tilde{a}_3 \dots, \tilde{a}_v)$  in a new  $\tilde{n}$  dimensional space such that

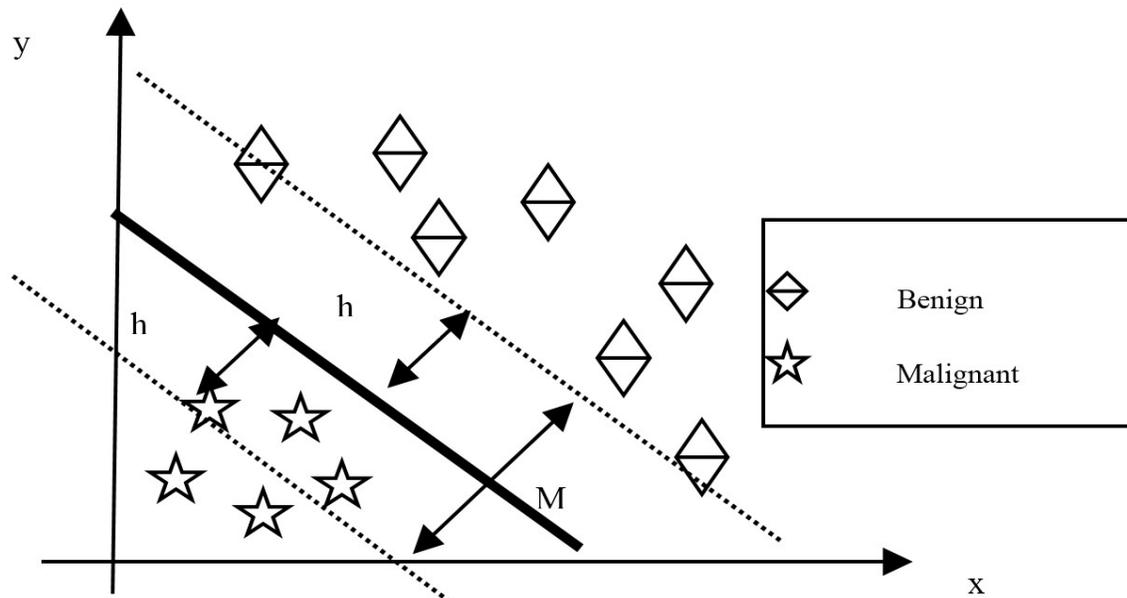
$$\tilde{a}_i = \sum_{l=1}^{\tilde{n}} c_{l,i} \lambda_l, \quad \tilde{n} \leq n \quad (10)$$

where the term  $c_{l,i}$  refers to the projection of the original vectors  $\tilde{a}_i$  on the eigenvectors  $\lambda_l$ .

The term  $\lambda_l$  is the eigenvectors corresponding to the biggest eigenvalues in  $\tilde{n}$ . The projection of  $c_{l,i}$  is known as the principal component of the original dataset. This projection explains the final steps in the PCA algorithm. The conditions of Equation (10) is both value of  $n$  and  $\tilde{n}$  must be greater than zero and the dimension size of  $\tilde{n}$  must be less than  $n$ . In this paper, 97% of the variance is utilized to select the most appropriate features and at the same time, most of the features information is also retain. The PCA also responsible to find the correlation between SURF features obtained in the previous phase. Since the SURF produce 3602 x 65 matrix, then the dimension reduction through PCA is crucial in this phase to increase the computational efficiency. Next, the classification phase is discussed to classify the selected SURF features of the breast masses either it is benign or malignant.

## 2.4. Classification

In this phase, the Support Vector Machine (SVM) is implemented because it has been proved as a good technique for the classification purpose [26]. The SVM hyperplane of two linear classes classification is illustrated as in Fig. 4.



**Fig.4.**The SVM hyperplane of two classes

From Fig.4, the hyperplane of the SVM is to separate two linearly classes. The margin  $M$  is twice the absolute value of the distance  $h$ . The points lying on the supporting planes is called the support vectors. The largest margin between two different classes is considered as the best hyperplane. The stars and diamonds shapes indicate two different classes. The wider margin can attain, hence the generalization ability is better. The two planes  $P_1$  and  $P_2$  for the closet point on one side and on the other side can be expressed as

$$\begin{aligned} P_1 &= \omega \circ x_i + h = 1 \\ P_2 &= \omega \circ x_i + h = -1 \end{aligned} \quad (11)$$

where  $\omega$  is an  $n$ -dimensional and the term  $h$  is the coefficient vector, normal to the hyperplane. Expanding the separate margin is equivalent to maximize the distance between hyperplane  $P_1$  and  $P_2$ . Hence, maximize margin is given by

$$m = \frac{2}{\|\omega\|} \quad (12)$$

such that

$$y_i(x_i \circ \omega + h) - 1 \geq 0, \quad \forall_i \quad (13)$$

The Equation (13) is also known as a convex quadratic problem. The Lagrange theory characterizes the solution of an optimization problem. The primal Lagrange function for Equation (13) is we want to minimize

$$LP = 0.5 * \|\omega\|^2 - \sum_{i=1}^n \Phi_i [y_i(x_i \cdot \omega + b) - 1] \quad (14)$$

$$LP = 0.5 * \|\omega\|^2 - \sum_{i=1}^n \Phi_i y_i(x_i \cdot \omega + b) + \sum_{i=1}^n \Phi_i \quad (15)$$

where the term  $\Phi_i$  represent the Lagrange multipliers and it must be positive for all value of  $i$ . Then, we differentiate with respect to  $\omega$  and  $b$  yields

$$\frac{\partial LP}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^n \Phi_i y_i x_i \quad (16)$$

$$\frac{\partial LP}{\partial b} = 0 \Rightarrow \sum_{i=1}^n \Phi_i y_i = 0 \quad (17)$$

After that, we replace Equation (14) and (16) into (17) and the optimization problem is obtained such as

$$\sum_{i=1}^n \Phi_i - 0.5 * \sum_{i=1}^n \sum_{j=1}^n \Phi_i \Phi_j y_i y_j x_i x_j \quad (18)$$

so that

$$\sum_{i=1}^n \Phi_i y_i = 0, \quad \Phi_i \geq 0 \quad \forall \Phi_i \quad (19)$$

The quadratic optimization problem with linear constraints is obtained. The quadratic algorithm are able to identify which training points  $x_i$  are support vectors with non-zero Lagrange multipliers  $\Phi_i$ . These supporting vectors is utilized to identify the decision functions.

Thus, the solution is

$$\omega = \sum_{i=1}^n \Phi_i y_i x_i = \sum_{i \in SVM} \Phi_i y_i x_i \quad (20)$$

and  $b = y_k - \omega x_k$  for any  $x_k$  such that  $\Phi_k \neq 0$  where  $x_k$  is the support vector and the term

*SVM* refers to the number of the support vector. Finally, the classification function is represent as

$$c(x) = \sum_{i \in SVM}^n \Phi_i y_i x_i x_j + b \quad (21)$$

There are various kernels that can be utilized in SVM. In this paper, the Fine-Gaussian kernel is utilized as the kernel for the classifiers since it has the properties of symmetric orientation, similar behaviour of the SURF features extraction method.

In addition to that, the Linear Discriminant Analysis (LDA) and Decision Tree (DT) classifiers are utilized to compare the performance of the proposed CADx. The LDA is utilized in this research since it is also considered as a practical tool for two classes classification [27]. The classification of LDA is obtained by reducing the class distance and expanding the class distance concurrently. The Decision Tree classifiers is also implemented in this research for comparison purpose. The DT is one of the data mining method that utilize a tree as a models with convergent solution [28]. The leaves refers to the classification and the branches represents the data in the features selections in this research.

### 3. RESULTS AND DISCUSSION

The following equations are the standard benchmark of a CADx performance measurement

$$A_c = \frac{TP + TN}{TN + FN + TP + FP} \quad (22)$$

$$FNR = \frac{FN}{FN + TP} \quad (23)$$

$$S_n = \frac{TP}{TP + FN} \quad (24)$$

$$S_p = 1 - S_n = \frac{TN}{TN + FP} \quad (25)$$

where  $A_c$  represent the accuracy of the proposed CADx . The  $FNR$  is the false negative rate.

In Equation (24) and (25), the  $S_n$  represent the sensitivity and  $S_p$  is the specificity respectively. The term  $TP, TN, FP$  and  $FN$  refers to true positive, true negative, false positive and false negative respectively. In this research, the false positive implies that the cancer is

benign and the biopsy procedure is needed for the particular patients. The proposed integrated PCA-SVM with SURF features extraction produces the following results as in Table 1.

**Table 1.** Results of the Proposed CADx

<b>Classifiers</b>	<b>Accuracy (%)</b>	<b>Az</b>	<b>FPR(%)</b>	<b>FNR(%)</b>	<b>Time (Sec)</b>
<b>SVM</b>	<b>87.5</b>	<b>0.96</b>	<b>12.3</b>	<b>0.17</b>	<b>2.3296</b>
LDA	55.5	0.55	1.12	0.87	2.3691
DT	57.2	0.52	1.05	0.94	5.4987
<b>PCA+SVM</b>	<b>92.9</b>	<b>0.98</b>	<b>7.1</b>	<b>0.01</b>	<b>2.1796</b>
PCA+LDA	57.9	0.56	1.1	0.9	2.8751
PCA+DT	57.6	0.54	1.49	0.63	1.2667

From Table 1, the table is divided into two main rows. The first three rows represent all classifiers (SVM, LDA and DT) without PCA component. Meanwhile, the PCA component is integrated with all classifiers represents in the last three rows.

In the first three rows, the SVM classifiers shows the highest accuracy with 87.5% compared to LDA and DT classifiers. In addition to that, the PFR and FNR value is 12.3% and 0.17% which indicate small error of misclassified is obtained. This FNR value also represent the smallest value achieved compared to all classifiers. In terms of accuracy, all processing time of classification is recorded in seconds. It can be observed that the SVM achieve the highest efficiency with 2.3296 seconds.

Focusing on the last three rows, the PCA+SVM classifiers shows an increasing in the accuracy with 4.4%, as well as the FNR from 0.17% to 0.01% compare to others classifiers. Moreover, the FPR is also reduced by 42.28%. The efficiency of the proposed method is also increasing with different of 0.15 seconds. This increasing is essentials in minimizing the computational costs issues.

In selecting the optimal kernels of SVM, four types of SVM kernel is compared in this research. These kernels are Quadratic, Cubic, Fine Gaussian and Medium Gaussian. Results obtained for kernel's selection are tabulated as in Table 2.

**Table 2.** Comparison of SVM's Kernels

<b>Kernels</b>	<b>Accuracy (%)</b>	<b>Az</b>	<b>FPR(%)</b>	<b>FNR(%)</b>	<b>Time (Sec)</b>
Quadratic	60.8	0.63	1.53	0.65	6.9985
Cubic	78.1	0.85	3.61	0.28	20.695
<b>Fine Gaussian</b>	<b>91.9</b>	<b>0.98</b>	<b>39.98</b>	<b>0.12</b>	<b>2.5581</b>
Medium Gaussian	56.1	0.66	5.59	0.49	6.3255
PCA+Quadratic	73.7	0.80	3.0	0.36	4.3529
PCA+Cubic	88.5	0.94	9.14	0.14	4.8208
<b>PCA+Fine Gaussian</b>	<b>92.9</b>	<b>0.99</b>	<b>20.25</b>	<b>0.11</b>	<b>2.5223</b>
PCA+Medium Gaussian	74.9	0.86	4.87	0.34	2.2233

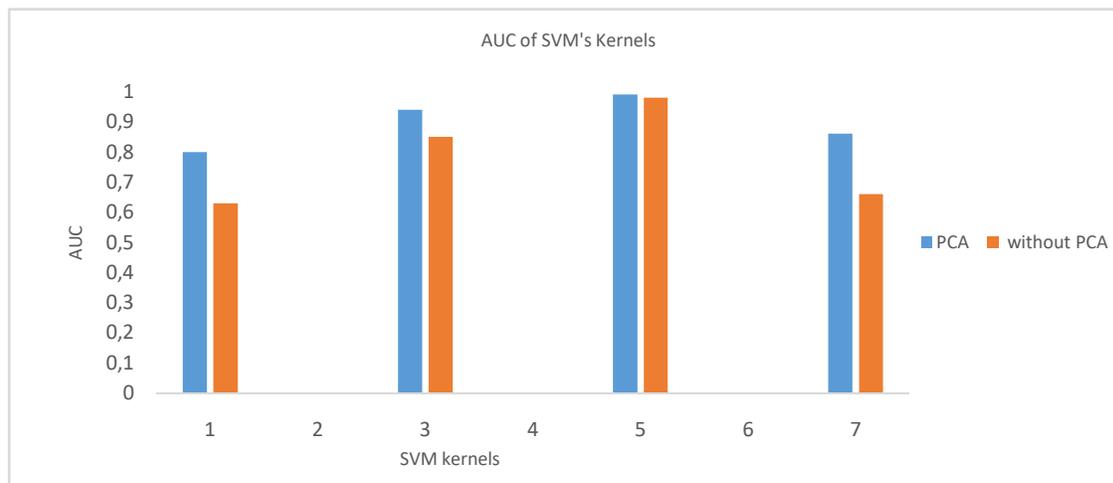
From Table 2, the experiment is conducted with two categories; with PCA component and without the PCA component. The first four rows indicates all kernels without PCA component and the last four rows represent all kernels with the PCA component.

In the first four rows, the Fine Gaussian kernel shows the highest accuracy with 91.9% among all kernels. This is followed by the Cubic kernel with 78.1%. Meanwhile, the Quadratic kernel is 60.8% and the Medium Gaussian kernel is 56.1%. Similarly, for the accuracy, the Fine Gaussian kernel achieve the highest efficiency time with 2.5581 seconds compared to other kernels.

It can be observed that there are significant increments in few perspectives in the last four rows. The PCA component integrated with Fine Gaussian kernels shows an increment of 0.1 and 0.01 for its accuracy and Az respectively. Moreover, there are 49.35% and 8.33% decreasing in the FPR and FNR respectively. For the accuracy, it can be observed that all kernels is also increasing significantly. The efficiency of the PCA+Fine Gaussian kernel is 2.5223 seconds. The PCA+Cubic kernels shows the second highest accuracy with 88.5% compared to PCA+Quadratic and PCA+Medium Gaussian. This is followed by PCA+Medium Gaussian kernel with increment of 18.8% accuracy.

These kernel also shows the highest increments among all kernels. Finally, the PCA+Quadratic kernel increase by 12.9% compared to the previous Quadratic kernel only. It can be concluded that the PCA component increase the performance of the accuracy and efficiency of all kernels.

The Az component is refers to the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) Curve. The AUC indicate that as the graph approach to 1, it is considered as a good or high accuracy. Fig. 5 illustrates the values obtained of all kernels conducted in this research.



**Fig.5.**The AUC of all kernels

From Fig. 5, there are four pair of graphs of the kernels with and without PCA component. The first graph represent the Quadratic kernel. This is followed by the Cubic and Fine Gaussian kernel and the last graph represent the Medium Gaussian kernel. Basically, all graphs show an increment after integrated with PCA component. We will focusing on third pair of graph since it produce the highest accuracy. The Fine Gaussian kernel without PCA is 0.98% and 0.99% with PCA component. Based on these findings, the Fine Gaussian kernel is utilized as the optimal kernel to the SVM classifiers. The performance of the proposed CADx is also compared with the existing CADx proposed by other researcher as tabulated as in Table 3.

**Table 3.** Comparison of the proposed CADx with the existing CADx

Reference	Year	Database	Feature	Feature	Accuracy (%)
			Extraction	Selection	
<b>This Research</b>	<b>2017</b>	<b>MIAS</b>	<b>SURF</b>	<b>PCA</b>	<b>92.9</b>
[15]	2016	MIAS	Haralick, Invariant	-	90.5
[29]	2012	DDSM	Zernike and Krawtchouk	-	90.2 81.0
[30]	2009	MIAS	Tolerant features	-	81.0
[31]	2000	MIAS	SI, FC, compactness	-	81.5

From Table 3, the current research conducted by [15] shows the accuracy of 90.5% of their proposed CADx for two classes classification. However, the feature selection component is not included in their study. This component might increase the accuracy of the CADx as well as to reduce the dimensionality of the features obtained. It can be observed that the proposed CADx achieve the highest accuracy with 92.9% compared to other listed researcher for the past seven years.

#### 4. CONCLUSION

In this paper, a novel CADx component based on SURF with PCA-SVM have been proposed to classify the breast cancer either it is benign or malignant. The SURF feature extraction method is utilize as a feature extraction. It is also proven that the PCA component in the features selection phase is able to reduce the FNR, as well as to increase the accuracy of the proposed CADx. Few kernels were experimented to find the optimal kernel for the SVM. The Fine Gaussian kernels shows the optimal kernel to be utilized in this research. Finally, the integration between PCA and SVM shows the highest accuracy among other classifiers to classify the cancer either benign or malignant. It is hope that this findings will slightly assist radiologist in making decision and to reduce the error rate at the diagnostic mammography stage. For the recommendation, it is recommended that the whole CADx is extended to an automate process. This is important in reducing the computational costs as well as to increase the performance of the CADx.

---

## 5. ACKNOWLEDGEMENTS

This research is supported by Ministry of Higher Education (MOHE) Malaysia. The author(s) would like to acknowledge MOHE for the supports and contributions.

## 6. REFERENCES

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA: A Cancer Journal for Clinicians, 2015, 65(1):5-29
- [2] DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. CA: A Cancer Journal for Clinicians, 2014, 64(1):52-62
- [3] Manoharan C, Lakshmi NS. Classification of micro calcifications in mammogram using combined feature set with SVM. International Journal of Computer Applications, 2010, 11(10):30-34
- [4] Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: Cross sectional survey. British Medical Journal, 2000, 320(7250):1635-1640
- [5] Li M, Zhou ZH. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2007, 37(6):1088-1098
- [6] Bassett LW, Gold RH. The evolution of mammography. American Journal of Roentgenology, 1988, 150(3):493-498
- [7] Tahmasbi A, Saki F, Shokouhi SB. Mass diagnosis in mammography images using novel FTRD features. In 17th IEEE Iranian Conference of Biomedical Engineering, 2010, pp. 1-5
- [8] Narváez F, Romero E. Breast mass classification using orthogonal moments. In International Workshop on Digital Mammography, 2012, pp. 64-71
- [9] Oliver A, Torrent A, Lladó X, Tortajada M, Tortajada L, Sentís M, Freixenet J, Zwigelaar R. Automatic microcalcification and cluster detection for digital and digitised mammograms. Knowledge-Based Systems, 2012, 28:68-75
- [10] Deserno TM, Soiron M, de Oliveira JE, Araújo AD. Computer-aided diagnostics of screening mammography using content-based image retrieval. In Proceedings of Society of Photo-Optical Instrumentation Engineers, 2012, pp. 1-9

- 
- [11] Azizi N, Zemmal N, Guiassa YT, Farah N. Kernel based classifiers fusion with features diversity for breast masses classification. In 8th IEEE International Workshop on Systems, Signal Processing and their Applications, 2013, pp. 116-121
- [12] Beura S, Majhi B, Dash R. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing*, 2015, 154:1-4
- [13] Rouhi R, Jafari M, Kasaei S, Keshavarzian P. Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Systems with Applications*, 2015, 42(3):990-1002
- [14] Pérez NP, López MA, Silva A, Ramos I. Improving the Mann-Whitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography. *Artificial Intelligence in Medicine*, 2015, 63(1):19-31
- [15] Yasiran SS, Salleh S, Mahmud R. Haralick texture and invariant moments features for breast cancer classification. *AIP Conference Proceedings*, 2016, 1750(1):1-6
- [16] Yasiran SS, Jumaat AK, Manaf M, Ibrahim A, Zarina WW, Malek A, Laham MF, Mahmud R. Comparison between GVF snake and ED snake in segmenting microcalcifications. In *IEEE International Conference on Computer Applications and Industrial Electronics*, 2011, pp. 597-601
- [17] Yasiran SS, Rahman WE, Ibrahim A, Kadir A, Mahmud R. Accuracy of enhanced distance active contour (EDAC) for microcalcifications segmentation. In *International Conference on Biomedical Engineering and Technology*, 2011, pp. 85-89
- [18] Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 2008, 110(3):346-359
- [19] Bhosale S B, Kayastha V S, Harpale V K. Feature extraction using surf algorithm for object recognition. *International Journal of Technical Research and Applications*, 2014, 2(4):197-199
- [20] Wang H, Oneata D, Verbeek J, Schmid C. A robust and efficient video representation for action recognition. *International Journal of Computer Vision*, 2016, 119(3):219-238
- [21] Li J, Wang Y, Wang Y. Visual tracking and learning using speeded up robust features. *Pattern Recognition Letters*, 2012, 33(16):2094-2101

- [22] Wen X, Shao L, Xue Y, Fang W. A rapid learning algorithm for vehicle classification. *Information Sciences*, 2015, 295:395-406
- [23] Viola P, Jones MJ. Robust real-time face detection. *International Journal of Computer Vision*, 2004, 57(2):137-154
- [24] Huang HF, Tai SC. Facial expression recognition using new feature extraction algorithm. *ELCVIA: Electronic Letters on Computer Vision and Image Analysis*, 2012, 11(1):41-54
- [25] Lüthi M, Blanc R, Albrecht T, Gass T, Goksel O, Büchler P, Kistler M, Bousleiman H, Reyes M, Cattin P, Vetter T. Statismo-A framework for PCA based statistical models. *Insight Journal*, 2012, 1:1-8
- [26] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995, 20(3):273-297
- [27] Yan Y, Ricci E, Subramanian R, Liu G, Sebe N. Multitask linear discriminant analysis for view invariant action recognition. *IEEE Transactions on Image Processing*, 2014, 23(12):5599-5611
- [28] Ardiansyah S, Majid MA, Zain JM. Knowledge of extraction from trained neural network by using decision tree. In *2nd IEEE International Conference on Science in Information Technology*, 2016, pp. 220-225
- [29] Narváez F, Romero E. Breast mass classification using orthogonal moments. In *International Workshop on Digital Mammography*, 2012, pp. 64-71
- [30] Rojas D A, Nandi AK. Development of tolerant features for characterization of masses in mammograms. *Computers in Biology and Medicine*, 2009, 39(8):678-688
- [31] Rangayyan RM, Mudigonda NR, Desautels JL. Boundary modelling and shape analysis methods for classification of mammographic masses. *Medical and Biological Engineering and Computing*, 2000, 38(5):487-496

**How to cite this article:**

Salleh S, Mahmud R, Rahman H, Yasiran S S. Speed up robust features (surf) with principal component analysis-support vector machine (pca-svm) for benign and malignant classifications. *J. Fundam. Appl. Sci.*, 2017, 9(5S), 624-643.