Mathematics Letters, eISNN: 2948-3735, 1 (2): 1-13, 2022

PERFORMANCE OF SUPPORT VECTOR MACHINE (SVM) USING BINARY ROBUST INVARIANT SCALABLE KEYPOINTS (BRISK) WITH DIFFERENT RBF KERNEL SCALES IN CLASSIFYING TYPES OF BREAST TISSUE

Siti Salmah Yasiran[1*] **and Suhaila Abd Halim**[2]

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor

[1*]salmahyasiran@uitm.edu.my, [2]suhaila889@uitm.edu.my

(indicates the corresponding author)*

ABSTRACT

Breast cancer is the major cause of death among women. The detection of breast cancer tissue is important for radiologists to diagnose breast cancer. Computer Aided Diagnosis (CADx) is one of the diagnosis systems to assist radiologists with an accurate diagnosis as well as to reduce the error rate. Hence, the objectives of the study are to propose a CADx for the classification of breast tissue. A set of 322 data was obtained from mini–Mammogram Image Analysis Society (MIAS) database used in the implementation of the CADx. The data is pre-processed first by using median filter and histogram stretching methods before segmented using hybrid Otsu and Mathematical Morphology. Then the features are detected and extracted by using the Binary Robust Invariant Scalable Keypoints (BRISK) features detectors. Next, the detected features are selected by using Principal Component Analysis (PCA) to reduce the dimensionality issue. Finally, the Support vector machine (SVM) is applied to classify the breast cancer tissue. In the classification stage, different values of Radial Basis Function (RBF) kernel scales are experimented concurrently with the box constraint levels. The accuracy of the proposed CADx is evaluated to measure its performance. The results demonstrate an accuracy of 83% is obtained which concludes that the proposed CADx has impressive performance.

Keywords*: Computer Aided Diagnosis (CADx), Support Vector Machine (SVM), breast cancer, Radial Basis Function (RBF) kernel, Binary Robust Invariant Scalable Keypoints (BRISK), Principal Component Analysis (PCA).*

Received for review: 02-06-2022; Accepted: 16-06-2022; Published: 01-11-2022

1. Background

Breast cancer was recorded the second highest new cases among Malaysian women in 2020 as compared to other cancers, colorectum, ovary, cervix uteri, and corpus uteri (Go & Form, 2020) The level of awareness among Malaysian women is still lack and it is sufficient to be identified to improve overall understanding on the disease. (Lee et al., 2019) mentioned that it is essential to focus on promotion in terms of risk factors and clinical presentations in early detection of breast cancer knowledge. There are different types of normal breast tissue that are fat, skin, glands, ducts and connective tissue (Malik et al., 2016) . The problem of manual breast tissue classification by radiologists is crucial and very time consuming since breast lesions images is composed of variety shape, texture, size and position (Maqsood et al., 2022).

According to Subashini et al., (2020), the breast tissue is an important indicator of breast cancer risk (Subashini et al., 2010). Hence, this issue motivates us to classify the types of breast tissue either it is fatty, dense, or glandular. The role of Computer Aided Diagnosis (CAD) is important to assist radiologist in reducing the error rate produced by human factors. The CAD system is divided into two subgroups: computer-aided detection (CADe) and computer-aided diagnosis (CADx). The CADe system was created to help radiologists discover and locate suspected areas of abnormality. The goal of the CADe is to help radiologists lower the rate of erroneous errors. The CADx system is intended for diagnosis and classification. It is also intended to detect and categorise benign and cancerous tissue. The goal of CADx is to help radiologists decide whether or not to perform a biopsy operation.

In general, the CAD component is composed of five major phases: data acquisition, pre processed, segmentation, features detection and extraction, features selection and finally the classification phase. According to Cheng et al.,(2006), the features detection and extraction phase is considered as a key steps since the performance of CADx depends so much more on the optimization features extraction and features selection (Cheng et al., 2006). There are diverse types of descriptors that have been utilized by researcher in various applications. This paper utilizes the binary robust invariant scalable keypoint (BRISK) descriptor in the features detection and features extraction phase.

The BRISK descriptor was originally introduced by Leutenegger et al. (2011). The purpose of BRISK is to overcome the weaknesses of speed up robust features (SURF) descriptor. Moreover, the BRISK descriptors are considered as an easy scalable for quicker implementation through lowering the number of sample points in the pattern. BRISK's special characteristics is essential for a many types of applications, particularly with strict real-time requirements or limited compute capacity (Leutenegger et al., 2011). Some applications on BRISK are in the fields of identification (Jang et al., 2020), video application (Khare et al., 2020), image stitching (Jatmiko & Prini, 2020) digital image forgery detection (Niyishaka & Bhagvati, 2020) and virtual-real registration (Xiang et al., 2020).

Focusing on the utilization of BRISK descriptors associated with mammogram images, Fanizzi et al., (2019) have proposed a novelty instrumentation for diagnosing of breast cancer using Contrast-Enhanced Spectral Mammography (CESM) images. They have proposed fully automatic CAD system tool to assist physicians. The segmentation phase is not included in their study and their classification is focusing on the benign and malignant classes. The BRISK descriptors are utilized together with SIFT, FAST, MSER and MinEigen descriptors for the features extraction. The Random Forest classifiers is utilized for the classification. It is found that the sensitivity and specificity of their proposed CAD is 87.5% and 91.7% respectively. (Fanizzi et al., 2019).

Abdudawood et al., (2018) have successfully developed an automatic and precise CAD for abnormalities classification of mammogram images. They have utilized the Local Binary Pattern (LBP) descriptors. The data was obtained from DDSM database. In their study, the BRISK descriptor is utilized as comparison purpose with the SURF descriptors. Few classifiers are also utilized in the classification stage. However, they have stated that the segmentation phase is not conducted since they only refer to the previous coordinate location of the region of Interest (ROI) which have been established by the radiologist. The performance of the proposed CAD is above 90% compared with different models (Abudawood et al., 2018).

The lesion of mammogram images is also plays an essential part for radiologist. Focusing on this lesion issue, de Sá Oliveira et al., (2018) have developed a novel method to find the lesion of mammogram images. The data is obtained form the CBIS-DDSM image database. In their study, the BRISK descriptors features was analysed together with the SIFT, SURF and ORB descriptors. The segmentation phase and features selection phase is not included in their study. They have found that the SURF descriptoirs superior among other descriptor method. (de Sá Oliveira et al., 2018).

Murtaza et al., (2021) have developed an automated detection model using machine learning approach which focusing on digital histology images. In their study, they have utilized ten features extraction methods including the BRISK descriptors. For the features selection phase, an algorithm is created to find the smallest number of redundant features. However, in their study, they did not utilize the segmentation phase. They have been using seven types of different classifiers for the classification phase. It is found that the performance of the developed model is 93% accurate (Murtaza et al., 2021).

Based on the reviews of BRISK in mammogram application, it is found that the BRISK descriptor is commonly been utilized for comparison purpose with other descriptors. There is no specific application on the BRISK descriptors itself at the features extraction stage. Thus, the objective of this study is to utilize the BRISK descriptors for both features detection and features extraction in the corresponding component of the proposed CADx. We utilize the advantage of the BRISK descriptors since it produces faster implementation by reducing the number of sample points in the pattern since mammogram is composed of complex white patches images. Then, the features will be selected through features selection stage. After that, the selected features will be classified by using SVM classifiers with different RBF kernel scales and box constraints levels simultaneously. Finally, the performance of the proposed CADx will be measured. The overall idea of the proposed CADx will be discussed in the next section.

2. Proposed Method

Figure 1 illustrates the proposed CADx to classify the types of breast tissue.

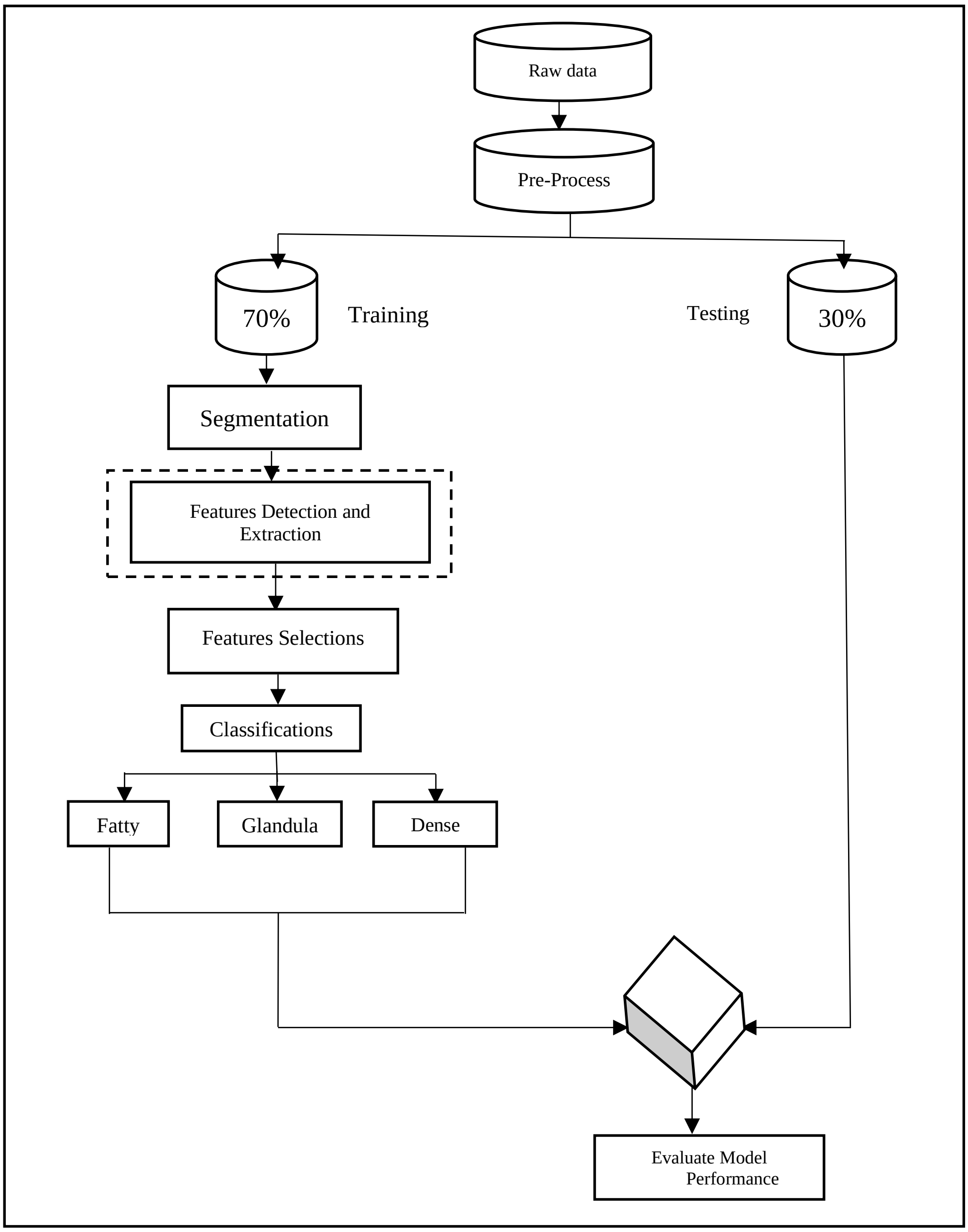

Figure 1. The proposed CADx.

From Figure 1, the proposed CADx can be divided into five main stages. The first stage is raw data obtained from the mini Mammogram Image Analysis Society (MIAS) database (Clark, 2012). The database is composed of 322 digitized films. It also includes the radiologist's truth values on the locations of any abnormalities that may be present. The database is composed of 200-micron pixel edge and all images are standardized to 1024 by1024 pixels. There are 109 fatty, 104 glandular and 117 dense. The remaining stages will be discussed in the next subsections.

2.1 Pre-Processed

The image is cropped based on the truth values provided by the database. Each image is resized to 255 by 255 pixels. The image enhancement method will be utilized to remove the noise as well as to improve the Region of Interest (ROI) image. The combination method of median filter and histogram stretching methods proposed by (S. Yasiran et al., 2011) will be utilized due to its strength which can increase the image contrast (Gonzalez et al., 2009).

2.2 Segmentation

2.2.1 Otsu and Mathematical Morphology

After finishing the pre-process stage, 70% of the data will be randomly selected to be trained and the remaining 30% of the data will be used later for the testing phase of the proposed CADx (Stone, 1974). In this stage, the hybrid method of Otsu and Mathematical Morphology proposed by (S. S. Yasiran et al., 2020) will be utilized for the segmentation method. Breast tissue images for each type of category will be segmented before proceeding to the next stage.

2.3 Features Detection and Features Extraction

Leutenegger et al. (2011) introduced the binary robust invariant scalable keypoint (BRISK) method primarily for real-time applications. The BRISK algorithm works by creating a scale-space pyramid, which is referred to as octaves and intra-octaves, and it finds corners using the AGAST algorithm (Mair et al., 2010). Next, it computes the FAST score for each scale space while searching for maxima to reduce redundancy. The BRISK descriptor archives rotation and orientation invariance by identifying the local gradient for each corner. For the illumination invariance, it compares the pixel-to-pixel intensity, obtains the results by testing the brightness, and a binary string of the descriptor is constructed. In this stage, all feature values from the segmented images will be detected and extracted by using BRISK. The BRISK process can be composed into three main phases which are: keypoint detection, keypoint description and descriptor matching. Each phase will be discussed on the next subsections.

2.3.1 Keypoint Detection

The keypoint detection of BRISK is motivated by Adaptive and Generic Accelerated Segment Test (AGAST) (Mair et al., 2010). Firstly, a scale space pyramid is constructed. This is followed by extraction of stable extreme points of sub-pixel thru AGAST. After that, the greyscale correlation of the arbitrary point pairs in the local image neighbourhood produces the binary features descriptors. Finally, the feature matching phase will be executed by using the Hamming distance method.

2.3.2 Keypoint Description

The keypoint description has a considerable impact on the future efficiency of descriptor matching, as well as the algorithm's overall performance. SIFT's keypoints have 128-vector descriptors, while SURF's keypoints have 64-vector descriptors. Unlike SIFT and SURF, the BRISK descriptor is represented as a binary bitstring (Mair et al., 2010). The sampling locations are indicated by little blue circles in Figure 2.

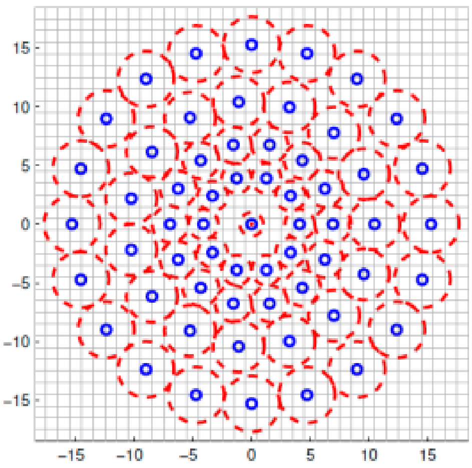

Figure 2. BRISK sampling pattern (Leutenegger et al., 2011).

From Figure 2, the small blue circles represent the location of the pixels to be considered. The red dashed circles present the smoothed area. The 40 by 40 pixels block centred on the interest point and four concentric circles are formed with N = 60 points. The distribution is uniform and equal spacing are obtained on each concentric circle (Leutenegger et al., 2011).

2.3.3 Descriptor Matching

The similarities between two feature points of the descriptors are examined and the descriptors are matched. The similarity of the descriptors is represented by computing the Hamming distance of the descriptor, since the BRISK technique employs a binary bit string of 1 and 0 to describe the extracted feature points. The Hamming distance is computed using XOR operation, which means that two values are involved in the computation. The result is 0. if their corresponding bits are the same, otherwise the value will be set to 1. The statistics of 1 are then counted, with the higher the number of 1 the greater the dissimilarity between the two descriptors, and vice versa. The mathematical formulation of the BRISK is represented as Equation (1).

$$a = a_1 a_2 \dots a_i \dots a_N \text{ and } b = b_1 b_2 \dots b_i \dots b_N \tag{1}$$

where a and b are two BRISK descriptors respectively. The value of a_i and b_i is either 1 or 0. The Hamming Distance equation is in Equation (2).

$$H(a,b) = \sum_{i=1}^{N} a_i \oplus b_i = \sum_{i=1}^{n} B(a_i, b_i) \tag{2}$$

$$\text{where } B(a,b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

where the term $\mathcal{B}(a_i, b_i)$ represents the bit inequality. The degree of two BRISK descriptors matching is calculated by using the Hamming Distance. A binary string of 512 bits is created by 512 comparisons.

2.4 Features Selection

2.4.1 Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) is utilized due to its capability to reduce the dimensionality of the input features (Lüthi et al., 2012), as well as to increase the performance of the proposed CADx. Moreover, the PCA is also capable to overcome the redundancy issues that arise in this stage. In this paper, 95% of the variance is utilized to select the most suitable features and concurrently, most of the feature's information is preserved. In addition to that, the PCA is also responsible to find the correlation between BRISK features obtained in the previous phase. Since the BRISK features produce 4199 by 64 matrices, then the dimension reduction through PCA is necessary to speed up the performance of the CADx. Next, the classification stage is conducted to classify selected BRISK features of the breast tissue, whether it is fatty, dense, or glandular.

2.5 Classification

2.5.1 Support Vector Machine (SVM)

In this stage, the Support Vector Machine (SVM) is implemented because it has been proved as the best method for classification (Cortes & Vapnik, 1995). The Radial Basis Function (RBF) is utilized in this paper to obtain the best performance for the classification. According to Cortes & Vapink (1995), the ability of SVM is equal to the number of training points $\mathfrak{y}$ that the model can be separated into $2^{\mathfrak{y}}$ distinct labels. It is also related to the total training data. The dimension $\mathscr{g}$ will influence the generalized error, as it is bounded by $\|\mathfrak{S}\|$, where the term $\mathfrak{G}$ refers to the weight vector of the separating hyperplane and the radius of the smallest sphere $\mathscr{j}$ that contains all the training points. It can be represented mathematically as $\|\mathfrak{S}\|^2\mathfrak{h} < \mathscr{j}^2$. Hence, the error is represented as $\mathbb{E} = \mathbb{E}_{train} + \mathbb{E}_{error}$ where the term $\mathbb{E}_{train}$ and $\mathbb{E}_{error}$ represent the training error and generalized error respectively. The training model $\mathcal{T}$ is defined by; $\mathbb{E}_{train}[\mathcal{T}] = \frac{1}{\mathfrak{h}}\sum_{i=1}^{n} 0.5|y_i - \mathcal{T}(.)|$ and the lower bound is represented by $\mathbb{E}[\mathcal{T}] \leq \mathbb{E}_{train}[\mathcal{T}] + \left[\frac{1}{\mathfrak{y}}\,\mathscr{g}\left(ln\frac{2\mathfrak{h}}{\mathscr{g}}+1\right) - \left(ln\frac{\mathfrak{h}}{4}\right)\right]^{\frac{1}{2}}$ where the term $(1-\mathfrak{h})$ represents probability bound for any function with dimension $\mathscr{g}$ which depends on independent data distributions (Awad & Khanna, 2015).

Since SVM classifiers find function $\ell = w^T x + b$ that separates two classes with the highest margin. Let x_i be a set of points that fit separable cases Θ_A and Θ_B and the distance from any points from the hyperplane, $H = \frac{|\ell(x)|}{\|w\|}$ and if the value of $\ell(x) = 1$ then the close data points will be fitted to class Θ_A. Conversely, if the value of $\ell(x) = -1$, then the close data points will be fitted to class Θ_B. Hence, the margin can be expressed as Equation (3).

$$w^T x + b = 1 \forall x \in \Theta_A \text{ and } w^T x + b = -1 \forall x \in \Theta_B \tag{3}$$

The objective functions in Equation (3) need to be minimized since

$$\frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|} \tag{4}$$

and $f = 0.5\|w\|^2$, subject to the constraint $y_i(w^T x + b) > 0$ where $i = 1,2,3,\ldots N$. In dealing with optimization cases, the cost of the error function is increasing by adding to the constraints and multiplying by the Lagrange multipliers. This can be represented as in Equation (5).

$$J(w,b,\alpha) = 0.5w^T w - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1] \tag{5}$$

where the terms w and b are primal variables respectively and the term α_i refers to the Lagrange multipliers.

There are four Karush-Kuhn-Tucker (KKT) conditions that generalize the Lagrange multipliers (Kuhn & Tucker, 2014). From Equation (3), the first condition is Primal constraints where $-[y_i(w^T x_i + b) - 1] \le 0, \forall i \in N$ The second condition is called the dual constraints which can be represented as; $\alpha_i \ge 0 \forall i \in N$. This is followed by the next condition, which is called the complementarity slackness. It takes the positive sides equal to zero of the first conditions: $[y_i(w^T x_i + b) - 1] = 0, \forall i \in N$. The final condition is the gradient of Lagrangian: $J(w,b,\alpha) = \left[w - \sum_{i=1}^{n} \alpha_i y_i x_i - \sum_{i=1}^{n} \alpha_i y_i \right] = 0$ and hence, $w = \sum_{i=1}^{n} \alpha_i y_i x_i$ and $\sum_{i=1}^{n} \alpha_i y_i = 0.$ A slack variables, $\Im$ is proposed since there exists errors in classification of SVM objective function. Now, the problem in the first condition is a minimization of the objective function

$$J(w,b,\Im) = 0.5\|w\|^2 + C\sum_{i=1}^{n} \Im_i \tag{6}$$

which subject to both constraints:

$$y_i(w^T x_i + b) \ge 1 - \Im_i \text{ and } \Im_i \ge 0 \text{ where } \forall i \in N \tag{7}$$

The term *C* is called the box constraint level parameter that varies and depends on the performance of the proposed CADx. If the values of *C* keep on increasing, then the number of support vectors will be decreased, and the training time will be increased. In contrast, if the values of *C* keep on decreasing, then the number of support vectors will be increased, and training time will be decreased (*MATLAB*, 2018) . The dual form of SVM can be represented as in Equation (8).

$$m_\alpha \left(\sum_{i=1}^{n} \alpha_i - 0.5\sum_{i,j}^{n} \alpha_i, \alpha_j, y_i, y_j x_i, x_j \right) \tag{8}$$

which subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $0 \le \alpha_i \le C \forall i \in N$ respectively The term m_α represent the maximization of the dual form of SVM. If the misclassification penalty, *C* approaches infinity, then the solution may converge to the linear separable category. (Apostolidis-Afentoulis & Lioufi, 2015).

2.5.2 Radial Basis Function (RBF) Kernel SVM

A kernel is used to transform the data into a higher-dimensional space. Rather than solving higher order separating hypersurface, a linear hyperplane can be acquired to split the different classes in the classification process. The RBF kernel is utilized since it is nonlinear and maps samples into a higher dimensional space. Instead of a linear kernel, RBF kernel has the ability to handle the nonlinear case between class labels and attributes (Keerthi & Lin, 2003). In addition to that, RBF kernels offer a number of hyperparameters that effects the complexity of the selected model and has little numerical complexities (Cortes & Vapnik, 1995). A kernel must satisfy the Hermitian format and positive semidefinite matrix. Mercer's theorem also needed to be fulfilled. which translates into evaluating the kernel or Gram matrix such as in Equation (9).

$$\varsigma(x,r)=\sum_i \wp_i(x),\wp_i(r),\ \text{where } \wp_i(x)\in \text{ Hilbert space.} \tag{9}$$

which implies to Equation (10).

$$\iint \varsigma(x,r)\iota(x)\iota(r)\,dxdu \geq 0 \tag{10}$$

In this paper, the Gaussian Radial Basis Kernel (RBF) is utilized and hence the Gram matrix will become as in Equation (11).

$$\varsigma(x,r)=\exp\left(-\frac{\|x-r\|^2}{\sigma^2}\right) \tag{11}$$

The kernel scales are tested simultaneously with the box constraint level to observe the performance of the proposed CADx. As the number of scales increases, the kernel matrix computation becomes ineffective which will lower the performance of the CADx (Awad & Khanna, 2015). Since the classification of breast tissue involves more than two classes, thus the multiclassification objective functions construct ς binary SVM classifiers. According to Kressel (1999), a multiclass classification will construct $\frac{\varsigma(\varsigma-1)}{2}$ binary SVM to discriminate 2 classes only and only needs evaluation of $(\varsigma-1)$ SVM classifiers (Kreßel, 1999). Then, the multiclass objective function can be represented as in Equation (12).

$$w^T_{(y_i)}\ell(x_i)+b_{(y_i)} \geq w^T_v\ell(x_i)+b_v+2-\Im^v_i, \text{provided that } \Im^v_i \geq 0 \tag{12}$$

where $\ell(x_i)=\varsigma(x,r)=\ell(x_i).\ell(x_j)$ and the solution of SVM must fulfil all KKT's conditions that have been stated earlier. Finally, after completing the training phase, 30% of the data will be used for the testing phase. This is to verify that the proposed CADx model will perform accurately as in the testing phase.

3. Experimental Results

In this paper, the Intel (R) Core (TM) i5-10300H CPU @ 2.50 GHz with 8.00 Giga Byte and 64-bit operating system, x64-based processor is utilized for the whole proposed CADx. Table

1 shows the accuracy performance of SVM with different box constraints and kernel scales from 1 to 25.

Table 1. Accuracy Performance of SVM.

KERNEL SCALE	**BOX CONSTRAINT 1**	**BOX CONSTRAINT 2**	**BOX CONSTRAINT 3**	**BOX CONSTRAINT 4**
1	83	83	82.9	82.9
2	82.9	82.9	82.9	82.9
3	82.6	82.7	82.8	82.9
4	81.9	82.1	82.5	82.5
5	80.7	81.1	81.7	82
6	78.3	79.3	80.2	80.5
7	75.7	76.8	77.5	78.4
8	72.9	73.7	74.4	75.6
9	70.5	71.1	72.2	72.9
10	68.8	69	69.4	70.7
11	67	67	67.3	67.9
12	65.7	65.5	65.6	66
13	63.9	63.9	64	64
14	62.6	62.7	62.6	62.6
15	61.8	61.8	61.8	61.9
16	61.1	61.2	61.2	61.2
17	60.5	60.7	60.7	60.7
18	60.2	60.2	60.5	60.1
19	59.6	59.7	59.7	59.8
20	59.3	59.3	59.6	59.5
21	58.8	59	59	59.1
22	58.6	58.7	58.8	58.6
23	58.1	58.2	58.5	58.4
24	57.8	58.1	58.1	58.1
25	57.6	57.8	57.9	58

The number of kernel scales is experimented until 25 since it has already achieved 57.6 % which represent that the performance of the CADx is poor (CD et al., 2015). The box constraint is utilized until level 4 since it will reduced the number of support vectors and also raised the training time (*MATLAB*, 2018). From Table 1, the accuracy is at the highest for box constraints 1 and 2 with 83%. Based on Equation (6), as the box constraint level parameter keep on increasing, the number of support vectors will be decreased, and we found that it reduces the overall accuracy of our proposed CADx. In addition to that, the accuracy is kept on reducing and start to reach approximate at 60% when the kernel scale is utilized at 17. The performance of different RBF kernels scales illustrates as in Figure 3.

Figure 3 shows the graph of different RBF scales accuracy. for kernel scales sizes from 1 to 25.

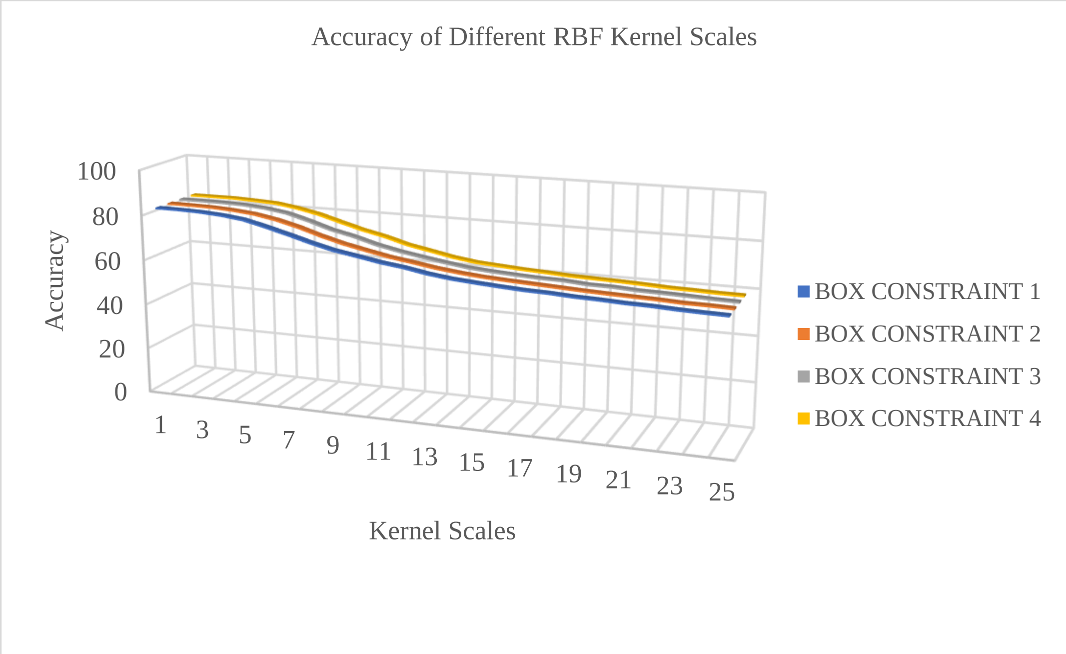

Figure 3. Graph of Accuracy for Different RBF Kernel Scale.

From Figure 3, it is clearly seen that the accuracy of the CADx is decreasing as the number of kernel scales increases for all box constraints (*MATLAB*, 2018). These results are also matched with the finding from Awad and Khanna (2015). The accuracy of the CADx was reduced as the number of kernel scales increases and leads to ineffective kernel matrix (Awad & Khanna, 2015).

4. Conclusion

This study utilizes the BRISK descriptors for both features detection and features extraction stage due to the advantages that have been mentioned earlier. The best overall accuracies for breast tissue classifications were achieved equal to 83% by using box constraints 1 and 2 for both kernels scales 1 and 2. On the other hand, the PCA could retain more than 95% of the entire whole features variance by using 45 features out of 64 features of the covariance matrices. The classifier and kernel function selection are necessary to get the best results of the proposed CADx. These findings may be developed to be a potential CADx for clinical assistant breast cancer diagnosis by providing the radiologists with an immediate second opinion.

For the recommendation, different types of features detection and features extraction may be further explored. The mix and match operations between both features detection and features extraction descriptors is also allowed (*MATLAB*, 2018). The operation can be conducted through various descriptors such as Scale Invariant Features Transform (SIFT), Speed Up Robust Features (SURF), Fast Retina Keypoints (FREAK) and Binary Robust Independent Elementary Features (BRIEF). As mentioned earlier, the features detection and features extraction stage are important in reducing the computational costs as well as to increase the performance of the CADx.

Acknowledgment

The authors would like to acknowledge the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Shah Alam for the support and contributions in completing this research.

References

Abudawood, T., Al-Qunaieer, F., & Alrshoud, S. (2018). An efficient abnormality classification for mammogram images. *2018 21st Saudi Computer Society National Computer Conference (NCC)*, 1–6.

Apostolidis-Afentoulis, V., & Lioufi, K.-I. (2015). SVM classification with linear and RBF kernels. *July): 0-7. Http://Www. Academia. Edu/13811676/SVM Classification with Linear and RBF Kernels.[21]*.

Awad, M., & Khanna, R. (2015). Support vector machines for classification. In *Efficient learning machines* (pp. 39–66). Springer.

CD, L., RD, W., DM, B., & al, et. (2015). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Internal Medicine*, *175*(11), 1828–1837. http://dx.doi.org/10.1001/jamainternmed.2015.5231

Cheng, H.-D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P., & Du, H. N. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, *39*(4), 646–668.

Clark, A. F. (2012). *The mini-MIAS database of mammograms*. http://peipa.essex.ac.uk/info/mias.html

Cortes, C., & Vapnik, V. (1995). Support-Vector Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297. http://doi.org/10.1023/A:1022627411411Networks. *Machine Learning*, *20*(3).

de Sá Oliveira, F. V., Alencar, G. A., & Cordeiro, F. R. (2018). Analysis of Keypoint Detection Algorithms for Mass Candidates Selection in Mammography Images. *Anais Do XV Encontro Nacional de Inteligência Artificial e Computacional*, 36–47.

Fanizzi, A., Losurdo, L., Basile, T. M. A., Bellotti, R., Bottigli, U., Delogu, P., Diacono, D., Didonna, V., Fausto, A., Lombardi, A., Lorusso, V., Massafra, R., Tangaro, S., & La Forgia, D. (2019). Fully Automated Support System for Diagnosis of Breast Cancer in Contrast-Enhanced Spectral Mammography Images. In *Journal of Clinical Medicine* (Vol. 8, Issue 6). https://doi.org/10.3390/jcm8060891

Go, W. T., & Form, P. A. (2020). Early Detection. *Lung Cancer*, *6*(7), 9–12.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Vol. 2. Digital Image Processing Using MATLAB*. Knoxville: Gatesmark Publishing.

Jang, D.-H., Kwon, K.-S., Kim, J.-K., Yang, K.-Y., & Kim, J.-B. (2020). Dog Identification Method Based on Muzzle Pattern Image. *Applied Sciences*, *10*(24), 8994.

Jatmiko, D. A., & Prini, S. U. (2020). Study and performance evaluation binary robust invariant scalable keypoints (BRISK) for underwater image stitching. *IOP Conference Series: Materials Science and Engineering*, *879*(1), 12111.

Keerthi, S. S., & Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, *15*(7), 1667–1689.

Khare, A., Mounika, B. R., & Khare, M. (2020). Keyframe extraction using binary robust invariant scalable keypoint features. *Twelfth International Conference on Machine Vision (ICMV 2019)*, *11433*, 1143308.

Kreßel, U. H.-G. (1999). 15 Pairwise Classification and Support Vector. *Advances in Kernel Methods: Support Vector Learning*, 255.

Kuhn, H. W., & Tucker, A. W. (2014). Nonlinear programming. In *Traces and emergence of nonlinear programming* (pp. 247–258). Springer.

Lee, M. S. (2019). Azmiyaty Amar Ma'Ruf C, Nadhirah Izhar DP, Nafisah Ishak S, Wan

Jamaluddin WS, Ya'acob SNM, et al. Awareness on breast cancer screening in Malaysia: a cross sectional study. *BioMedicine, 9*(3), 18.

Leutenegger, S., Chli, M., & Siegwart, R. (2011). BRISK: Binary Robust invariant scalable keypoints. In *Proceedings of the IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2011.6126542

Lüthi, M., Blanc, R., Albrecht, T., Gass, T., Goksel, O., Büchler, P., Kistler, M., Bousleiman, H., Reyes, M., & Cattin, P. (2012). Statismo-A framework for PCA based statistical models. *The Insight Journal*, *2012*, 1–18.

Mair, E., Hager, G. D., Burschka, D., Suppa, M., & Hirzinger, G. (2010). Adaptive and generic corner detection based on the accelerated segment test. *European Conference on Computer Vision*, 183–196.

Malik, B., Klock, J., Wiskin, J., & Lenox, M. (2016). Objective breast tissue image classification using Quantitative Transmission ultrasound tomography. *Scientific Reports*, *6*(1), 1–8.

Maqsood, S., Damasevicius, R., & Maskeliunas, R. (2022). TTCNN: A Breast Cancer Detection and Classification towards Computer-Aided Diagnosis Using Digital Mammography in Early Stages. *Applied Sciences*, *12*, 1–27. https://doi.org/10.3390/app12073273

MATLAB. (2018). Natick, Massachusetts: The MathWorks Inc. https://doi.org/9.7.0.1190202

Murtaza, G., Wahab, A. W. A., Raza, G., & Shuib, L. (2021). Breast Cancer Detection via Global and Local Features using Digital Histology Images. *Sukkur IBA Journal of Computing and Mathematical Sciences*, *5*(1), 1–36.

Niyishaka, P., & Bhagvati, C. (2020). Copy-move forgery detection using image blobs and BRISK feature. *Multimedia Tools and Applications*, *79*(35), 26045–26059.

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(2), 111–147. http://www.jstor.org/stable/2984809

Subashini, T. S., Ramalingam, V., & Palanivel, S. (2010). Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, *114*(1), 33–43.

Xiang, Z., Liyu, T., & Ding, L. I. N. (2020). Virtual-real registration method of natural features based on binary robust invariant scalable keypoints and speeded up robust features. *Journal of Computer Applications*, *40*(5), 1403.

Yasiran, S., Eny, W., Wan, Z., Abdul Rahman, A., Ibrahim, A., Jumaat, A., & Mahmud, R. (2011). Accuracy of Enhanced Distance Active Contour (EDAC) for Microcalcifications Segmentation. *2011 International Conference on Biomedical Engineering and Technology*.

Yasiran, S. S., Salleh, S., & Mahmud, R. (2020). Otsu and Mathematical Morphology for Breast Cancer Classification. *ASM Science Journal*. https://doi.org/10.32802/asmscj.2020.sm26(4.9)