Received: May 5, 2020. Revised: May 26, 2020.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Second Order Statistical Method for Breast Thermal Images Classification

Aris Marjuni[1]* **Oky Dwi Nurhayati[2]**

[1]*Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia*
[2]*Department of Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Indonesia*
* Corresponding author's Email: aris.marjuni@dsn.dinus.ac.id

Abstract: This paper is proposed to reveal the classification of randomized thermograms tabulated by the four features in the second-order statistic features extraction method involving the angular second-moment, contrast, correlation, and entropy values. The randomized thermograms, or breast thermal images, are captured by the thermal camera in the radiotherapy laboratory and analyzed using the mathematical method of measurement. All four features are used as input principal component analysis (PCA) to classify the types of thermograms after preprocessed using the wiener filtering and histogram equalization. Experimental results show that the method is quite promising to distinguish the thermal images by the sensitivity and specificity rates at 92% and 95%, respectively.

Keywords: Thermal images, Wiener filtering, Histogram equalization, Second-order statistic, Principal component analysis.

1. Introduction

Breast cancer is the most suffered cancer among women worldwide. WHO reported that about 2.1 million cases each year, and estimated 627,000 women died in 2018 due to breast cancer [1, 2]. These facts are significantly increased compared to the 2012 cases, where there were 1.7 million cases with 522,000 deaths [3]. In further reported that breast cancer killed more than 500,000 women on average around the world every year [4]. WHO also reported that breast cancer in the majority of women often detected in late stages because of the limited resources and low health system. The survival rates of women with breast cancer too low with ranging from 10 to 40%. To increase the survival rates, the two methods are recommended, which are early diagnosis and screening [1, 4-8]. Principally, when cancer can be detected at an early stage disease, it should save women's life [6].

The early diagnosis method concerns to improve the diagnosis service by providing timely access to cancer treatment. Whereas, the screening method concerns to identify and evaluate cancer early before any symptoms appear. These are some popular screening methods, such as mammography, clinical breast exam, and breast self-exam [1, 4]. Mammography currently is used as the standard tool for screening breast cancer. However, due to the inherent disadvantages in imaging dense breast tissues, the new alternative technique has been developed using thermography [7-9].

In the breast cancer field, thermography, which is also known as thermal imaging, is generally an imaging technique using an infrared camera to examine the temperature of the breast surface. Infrared thermography or thermogram is usually used in medical applications such as to identify the abnormal affected area to find the early malignant tumors and to assist in the treatment of breast cancer, different diabetes, peripheral neuropathy, and peripheral artery disorders [10]. Infrared thermography also has shown to be a promising technique for the early diagnosis of breast pathologies [6, 9]. Other studies related to thermography mostly discuss the comparison of sensitivity levels, a specificity of breast cancer screening methods with mammography, thermography, and ultrasonography [11].

DOI: 10.22266/ijies2020.0831.32

The thermogram is more proper screening and has a lower cost than other types of screening methods like the mammogram, ultrasound, and magnetic resonance imaging depending on the temperature of the breast and surrounding area by using a special heat-sensing camera to determine the heat in the region of breasts [12]. Meanwhile, the image processing carried out on the thermogram mostly uses statistical methods to calculate the degree of asymmetry between the left and right breast [13]. Nurhayati et al. [13] proposed the first-order statistical method for breast thermal image classification. They revealed the randomized thermograms classification which is tabulated by the first-order statistics method including the mean values, entropy values after the mathematical method of measurement. Combining these statistical parameters with a principal component analysis method results in a better analysis to distinguish the types of thermal images.

Thermogram, as biomedical or bioinformatics data, has commonly high dimensional in the number of attributes and/or record numbers [14-16]. Hence, it is necessary to select the most important attributes or features for classification using machine learning algorithms [15]. In breast cancer detection purposes, a principal component analysis (PCA) technique has widely used for feature selection and it is usually utilized with classification algorithms. Started with the PCA's advantage, this study is only focused on the use of PCA in breast cancer detection and analysis, which is motivated by some following related works.

Lyng et al. [3] proposed the combination of PCA with linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) methods for breast cancer diagnosis. However, both methods gave similar results in sensitivity and specificity performances. The sensitivity and specificity values are approximated by 83% and 80%, respectively. The PCA and LDA method for breast cancer diagnosis is also proposed by Moisoiu et al. [17] with the sensitivity and specificity of their proposed method are 81% and 95%, respectively. Jamal et al. [15] used PCA and $k-means$ clustering algorithm to predict breast cancer using Wisconsin Breast Cancer (WBC) datasets. The performance of PCA and $k-means$ combined with support vector machine (SVM) and XGBoost are compared in their study. The performance of PCA combined with SVM and XGBoost outperformed the $k-means$ on almost of their experiments.

Liu and Ma [16] utilized PCA with the SVM and backpropagation neural network (BPNN) algorithms to achieve breast cancer recognition effectively. The pathological images are extracted using PCA to get the most feature and then classified using SVM and BPNN. The combined PCA and SVM algorithms gave better accuracy, sensitivity, and specificity in breast cancer detection, rather than using the mixture of PCA and BPNN. The use of PCA and SVM is also proposed by Luna-Rosas et al. [18]. They proposed the PCA and parallel SVM to find the optimal response time in the automated breast cancer detection. They used the PCA method to find the distribution of two classes that are healthy and damaged tissue. Some classifiers were used to detect high specificity and sensibility rates and optimized by SVM to optimize the response time in automated breast detection.

Sahu et al. [19] proposed a hybrid method between the PCA with an artificial neural network (ANN) and random forest (RF) to classify breast cancer using WBC datasets. After dimensional reduction using the PCA, the datasets then classified by ANN and RF. In terms of sensitivity and specificity, the hybrid of PCA and ANN performed better than the PCA and RF. However, the hybrid of PCA and RF performed better in accuracy than the PCA and ANN. Dzulkalnine et al. [20] proposed the fuzzy PCA (FPCA) with SVM for breast cancer classification. The fuzzy PCA is adopted as a feature selection method to find the optimum significant factors for breast cancer detection. The hybrid of FPCA and SVM is compared to the benchmark methods, which are the SVM method only, and the hybrid of PCA and SVM. Experimental results show that the FPCA and SVM performed better in accuracy, sensitivity, specificity, and also AUC compared to the SVM and the hybrid of PCA and SVM.

Based on the advantage and popularity of the PCA for breast cancer detection and analysis, this study is also carried out to use the PCA to extract the thermogram features. In this paper, we propose the new approach to determine the thermogram classifications that are the normal thermogram, early cancer thermogram, and advanced cancer thermogram. For this classification purpose, the second-order statistical method is used to evaluate the important features from the thermogram image that affect breast cancer detection. The result of this study can be used to determine the condition of monitored cancer patients from the thermogram results. Thus, in this paper, the use of PCA and second-order statistical method is proposed to classify the breast thermal images for breast cancer detection.

This paper is presented in several sections, as follows. Section 2 presents an underlying theory related to the use of PCA in breast cancer analysis. Section 3 presents the proposed method using the second-order statistical method based on PCA. The pre-processing of the thermogram images using Wiener filtering and histogram equalization also presented in this section. Section 4 presents the experimental setup of this study including the dataset preparation, scope definition, and performance evaluation method. The experimental results and discussion are presented in section 5. Finally, the conclusions and future works of this study are summarized in section 6.

2. The underlying theory

PCA is a popular technique in statistical data analysis, feature extraction, and data reduction based on a singular value decomposition [21, 22]. It is commonly used to reduce the dimensionality of data to examine its underlying structure and the covariance or correlation structure of a set of variables. While singular value decomposition provides a simple means for identification of the principal components for classical PCA, solutions achieved in this manner may not possess certain desirable properties including robustness, smoothness, and sparsity [8, 11, 23, 24].

PCA is a deterministic method for reducing the dimensionality of a dataset by transforming some possibly correlated variables into a smaller number of linearly uncorrelated variables, which are called principal components [13, 23]. Technically, PCA involves the variances and covariances through linear combinations of the provenience variables without missing significant information of the source data [25]. The goal of the PCA is to reduce the variable space from a large set of variables into a smaller set of variables. These goals are achieved by maximizing the variance of projected data and minimizing the mean squared error between the data points and projected data. PCA of a large data space will produce some orthonormal basis vectors in the form of a collection of eigenvectors from a particular covariant matrix, which can optimally represent the distribution of data [26].

In the imaging technique, the form of eigenspace representation can be obtained by transforming the PCA into a set of images. The result of this transformation is an orthonormal basis vector which is used to form a sub-vector space called a feature space [22, 26]. The main advantages of PCA are mainly in computational aspects by reducing the complexity so that more efficient in capacity and memory requirements than the nonlinear computational. PCA is also reducing the redundancy of data through the orthonormal components and increasing the maximum variation so that it produces a low noise sensitivity. Once the patterns in the data have been found and the data and examples have been compressed by reducing the number of dimensions, the information in it would not be much in lost [13].

3. Proposed method

This research involves image processing methods to process thermograms. For pre-processing, this study uses a Wiener filtering and histogram equalization to improve the image quality of the breast thermogram, as illustrated in Fig. 1.

After the pre-processing image, the next step is to extract the features of an image. Feature extraction is the process of taking the characteristics found in an object in the image to recognize the object. Feature extraction is the first step in image classification and interpretation [27]. One method used in feature extraction is first-order statistical feature extraction and second-order feature extraction [13].

First-order feature extraction is a method of characterization based on image histogram characteristics. The histogram shows the probability of the appearance of the grey pixel degree values in an image [10, 11, 13, 28]. Second-order statistical feature extraction is done with a co-occurrence matrix, which is an intermediate matrix that represents correlations of pixels in the image in various orientations and spatial distance as describes in Fig. 2. Co-occurrence means joint events, i.e. the number of occurrences in one level of neighboring pixel value with another pixel value level within a certain distance (d) and angle orientation (θ). Distance is expressed in pixels and orientation is expressed in degrees. The orientation is formed in four angular directions with 45° angle intervals, namely 0°, 45°, 90°, and 135° while the distance between pixels is usually set at 1 pixel [27]. After obtaining the co-occurrence matrix, the second-order statistical characteristics that represent the observed image can be calculated.

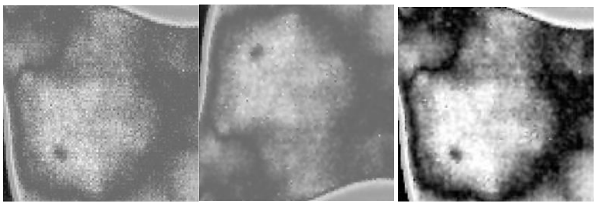
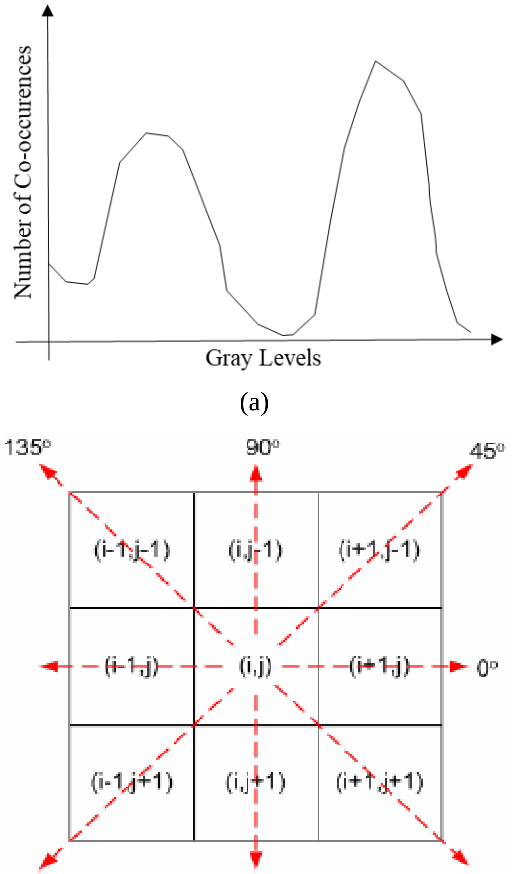

(a) (b) (c)

Figure. 1 (a) original thermogram, (b) thermogram after filtered wiener, and (c) after histogram equalization



(a)

(b)

Figure. 2 (a) Image histogram and (b) the relationship between neighboring pixels as a function of orientation and spatial distance [27]

Several second-order statistical features are calculated, namely angular second moment, contrast, correlation, and entropy [27].

a. Angular Second Moment (ASM) functions to show the characteristics of the image homogeneity that can be obtained by Eq. (1) where $p(i,j)$ represents the value in row i and column j in the co-occurrence matrix.

$$ASM = \sum_i \sum_i \{p(i,j)\}^2 \qquad (1)$$

b. Contrast (CON) functions to show the range of the spread (moment of inertia) elements of the image matrix. If it is located far from the main diagonal, the contrast value is large. Visually, the contrast value is a measure of variation in the grey degrees of an image range shown in Eq. (2).

$$CON = \sum_n n^2 \left\{ \sum_i \sum_{\substack{j \\ |i-j|=n}} p(i,j) \right\} \qquad (2)$$

c. Correlation (COR) functions to show the size of the linear dependence of the grey image degree so that it can provide a hint for the linear structure in the image that can be obtained in Eq. (3) with μ_x and μ_y are the average values of p_x and p_y, respectively. Whereas, σ_x and σ_y are and the standard deviation values of values of p_x and p_y, respectively.

$$COR = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_i\mu_j}{\sigma_x\sigma_y} \qquad (3)$$

d. Entropy (ENT) describes the value of shape irregularity. The entropy value is large for images with a uniformly greyed-out degree transition and is of little value for the irregular (varied) image structure obtained in Eq. (4).

$$ENT = -\sum_i \sum_j p(i,j)^2 \log(p(i,j)) \qquad (4)$$

4. Experimental setup

This study uses 170 breast thermogram images in 256×192 of size for the experiment. These breast thermogram images are categorized into the normal thermograms (50 images), early breast cancer thermograms (50 images), advanced breast cancer thermograms (50 images), and undergoing breast cancer thermograms (20 images). All of these thermogram images were taken from breast cancer patients using a thermal camera in the radiotherapy laboratory of the general hospital Dr. Sarjito, Indonesia. All of the patients of this study were only female patients, without any age restrictions, and the observed cancer location around in the breast area.

The scope of this experiment is specifically to simulate and evaluate a breast cancer detection method using thermogram images by applying a second-order statistical feature extraction. The features used in this proposed method are the ASM, CON, $CORR$, and ENT, which are formulated in Eq. (1) to Eq. (4), respectively. Besides, this study also uses the PCA method to calculate eigenvalues, eigenvectors, and covariance matrix of thermogram data images. The PCA method flow chart is presented in Fig. 3.

For evaluation, the performance of the second-order statistical method for breast thermal image classification in this study is measured by sensitivity, specificity, and receiver operating characteristics (ROC) curve. Sensitivity, also called true positive

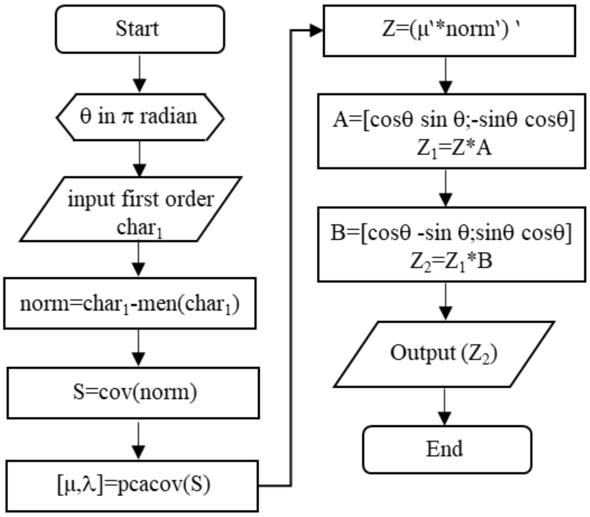

Start
θ in π radian
input first order char$_1$
norm=char$_1$-men(char$_1$)
S=cov(norm)
[μ,λ]=pcacov(S)
Z=(μ'*norm') '
A=[cosθ sin θ;-sinθ cosθ]
Z$_1$=Z*A
B=[cosθ -sin θ;sinθ cosθ]
Z$_2$=Z$_1$*B
Output (Z$_2$)
End

Figure. 3 The PCA method flow chart

rate (TPR), measures the percentage of actual positives or cancer detected that are correctly positives. The sensitivity value represents the probability of a positive result that the patient has the disease. Specificity (also called true negative rate) measures the percentage of actual negatives or no cancer detected that are correctly negatives.

The specificity value represents the probability of a negative result that the patient is healthy. The sensitivity and specificity values are computed using the following terminology. The diagnostic activities of this research can be seen as probability events as follows:

a. $T+$: a positive result or cancer detected.
b. $T-$: a negative result or cancer undetected.
c. $D+$: indicates diseases (positive).
d. $D-$: indicates no disease (negative).

The quality of a procedure or diagnostic test can be seen from the conditional probabilities below:

a. Sensitivity: $Sens = P(T+ \mid D+)$ (5)
b. Specificity: $Spec = P(T- \mid D-)$ (6)

A measurement or diagnostic test is considered to be good if it has high sensitivity and specificity value (close to 1). The plot of TPR versus the false positive rate (FPR or $1 - sens$) will generate the curve in the unit square, which called the ROC curve [23]. Graphically, the ROC curve represents the ability of the diagnostic tests to distinguish the "diseased" and "non-diseased" states [29].

5. Results and discussion

The results of this study are carried out with diagnostic tests of sensitivity and specificity parameters to determine whether the patient was

Table 1. The diagnostic activities from patients data

	$T-$	$T+$	Total patient
$D-$	74	4	78
$D+$	8	98	106
$D1-$	65	4	69
$D1+$	4	15	19
$D2-$	74	8	84
$D2+$	4	83	89

Note: $D1+, T+$ = early cancer; $D2+, T+$ = advanced cancer; $D-$ = no cancer detected

Table 2. Sensitivity and specificity based on score level

Positive test criteria	Sensitivity	Specificity	1-Specificity
1 ≥ score	1.00	0.00	1.00
1 ≤ score ≤ 2	0.79	0.94	0.06
3 ≤ score	0.93	0.88	0.12

Note: 1 = normal; 2 = early cancer; 3 = advanced cancer

Table 3. Average results of second-order statistical characteristic values

Type of thermogram	ASM	CON	COR	ENT
Normal	0.0024	103.10	0.97	11.001
Early	0.0012	141.86	0.94	11.410
Advanced	0.0010	145.04	0.96	11.627

diseased or not. The data distribution of the diagnostic activities is summarized in Table 1. The values of $D+$ and $T+$ are categorized as early cancer. The values of $D2+$ and $T+$ are categorized as advanced cancer, while the $D-$ values are categorized as no cancer detected.

The diagnostic activities values in Table 1 are then used to compute the sensitivity and specificity that are formulated in Eq. (5) and Eq. (6), respectively. The sensitivity and specificity of the diagnostic score level are summarized in Table 2, where there are three levels, that are: normal, early cancer, and advanced cancer criteria.

The four second-order statistical features extracted from each thermogram are the ASM, CON, $CORR$, and ENT. The average results of the second-order statistical characteristic values are presented in Table 3.

Using the PCA method, the eigenvectors and the eigenvalues of the second-order statistical features are shown in Table 4. The ASM plot characteristics in the correlations of all raw thermogram data for normal, early cancer, and advanced cancer are shown in Fig. 4.

Table 4. The eigenvectors and eigenvalues range of the second-order statistical features

Features	Eigenvectors			Eigenvalues		
	Normal	Early	Advanced	Normal	Early	Advanced
$ASM; CON$	[-1,0]; [0,1]	[0,1]; [1,0]	[0,-1]; [-1,0]	[2124; 0]	[2470; 0]	[2124; 0]
$ASM; CORR$	[-1,-0.001]; [-0.001,1]	[0.05,0.99]; [0.99,-0.05]	[0,1]; [1,0]	[0.002; 0]	[0; 0]	[0.002; 0]
$ASM; ENT$	[-1,-0.003]; [-0.003,1]	[-0.002,1]; [1, 0.002]	[0,-1]; [-1,0]	[0.767; 0]	[1.07; 0]	[0.767; 0]
$CON; CORR$	[-0.0003,-1]; [-1,0.0003]	[-1,0.0002]; [0.0002,1]	[0,1]; [1,0]	[2124; 0]	[2470; 0]	[2124; 0]
$CON; ENT$	[-0.01,-1]; [-1,0.01]	[-0.99,-0.02]; [-0.02,0.99]	[0,-1]; [-1,0]	[2124; 0.6]	[2470; 0.3]	[2124; 0.6]
$CORR; ENT$	[-1,0.008]; [0.008,1]	[-0.009,1]; [1,0.009]	[0,-1]; [-1,0]	[0.767; 0.002]	[1.07; 0]	[0.767; 0.002]

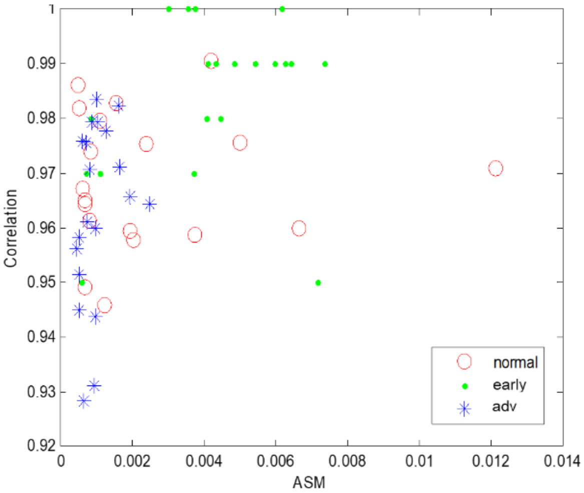


Figure. 4 ASM plot characteristics result in the correlations of all raw thermogram

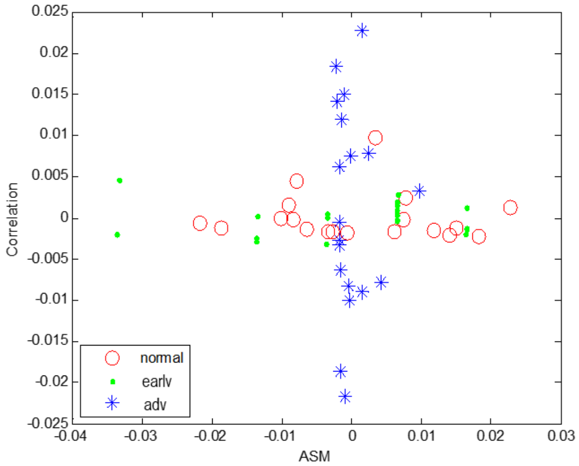


Figure. 5 The plot of the normalization thermogram data of the ASM feature towards the $CORR$

The plot of the normalization thermogram data of the ASM feature towards the correlation feature is displayed in Fig. 5. The decoupling results from the transformation of the matrix $A = [\cos\theta \sin\theta ; -\sin\theta \cos\theta]$ and the characteristics of the thermogram data correlation are shown in Fig. 6. Whilst the decoupling results of the matrix $B = [\cos\theta - \sin\theta ; \sin\theta \cos\theta]$ and the characteristics of the thermogram data correlation are presented in Fig. 7.

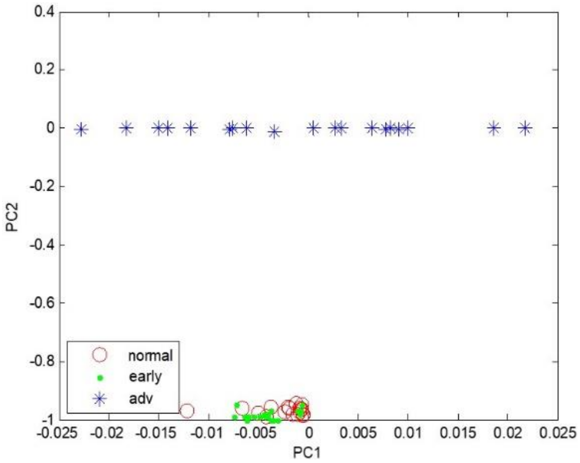


Figure. 6 The decoupling results of the first transformation at angle A characterize ASM vs $CORR$

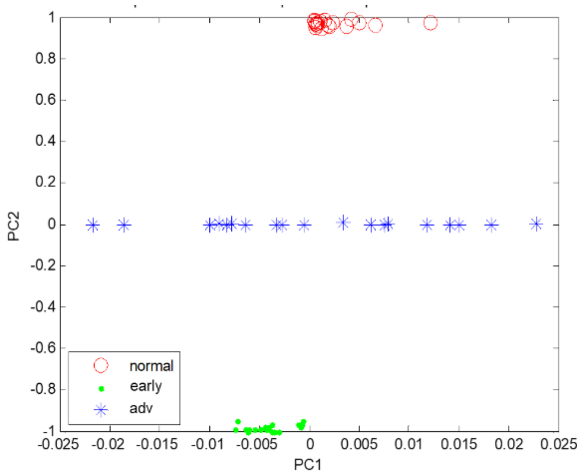


Figure. 7 The further decoupling results of the angle B characterize ASM vs $CORR$

Received: May 5, 2020. Revised: May 26, 2020.

The further decoupling results of the angle B of the ASM characteristic pair to the correlation feature, as shown in Fig. 7, give significant results in separating the three types of thermograms, namely normal thermogram, early cancer thermogram, and advanced cancer thermogram. As shown in Fig. 7, the normal thermograms have values close to 1, the early cancer thermograms have values close to -1, and the advanced cancer thermograms have values close to zero (0).

The plot for contrast characteristics results towards the correlation characteristics after decoupling due to the transformation of the trigonometric matrix B is presented in Fig. 8. It provides significant information to the separability of the thermogram gives a much different range. The normal thermograms have values close to 1, the early cancer thermograms have values close to -1, and the advanced cancer thermograms have values close to zero (0).

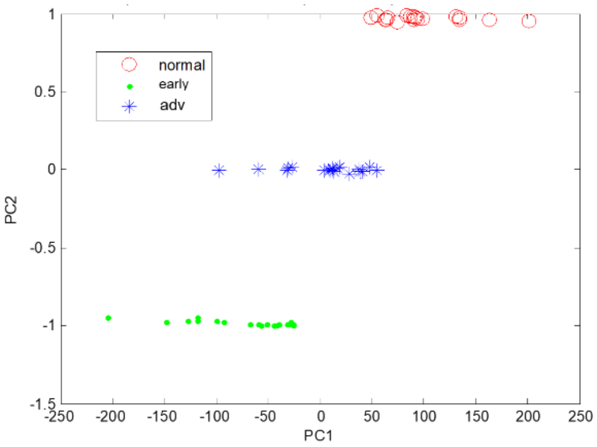


Figure. 8 The decoupling results due to the transformation of angle B on the contrast characteristic towards correlation

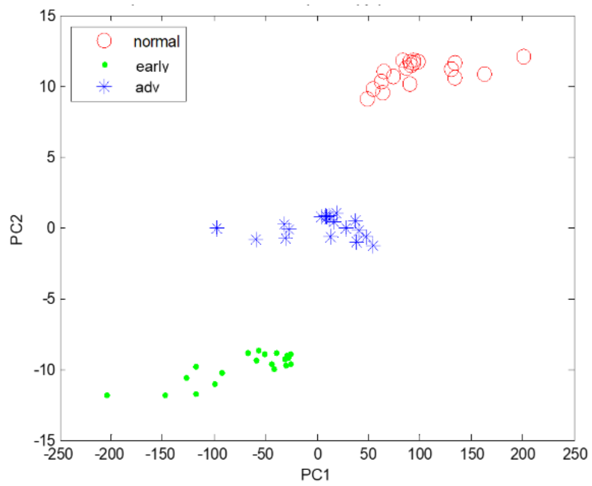


Figure. 9 The further decoupling results in angle B of the contrast characteristic towards entropy

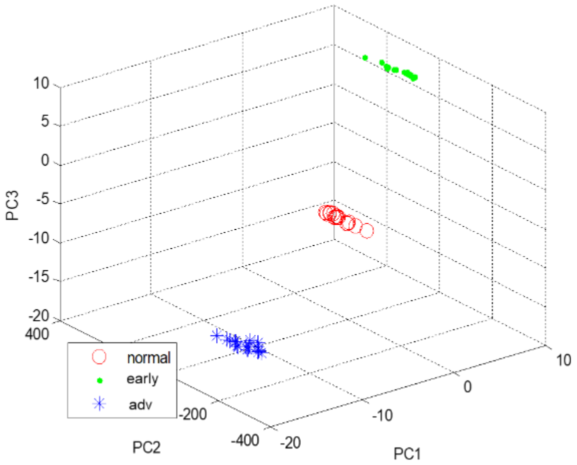


Figure. 10 The decoupling results of the B trigonometric matrix ASM characteristic towards contrast and correlation

Similarly, the plot of the contrast towards the entropy after further decoupling at angle B is shown in Fig. 9. It shows the further decoupling results of the trigonometric matrix B in contrast and entropy characteristics. The three types of thermograms are well separated so that there is no information redundancy among the three types of thermograms. The normal thermograms have a range of values 5 to 15, the early cancer thermograms have a range of values -5 to -15, and the advanced cancer thermograms have a range of values -4 to 4.

Some of the second-order statistical characteristics result in three dimensions on ASM, CON, $CORR$, and ENT features are divided into two groups, namely the classification results with strong correlation and the classification results with weak correlation. The further decoupling results using eigenvalues on the ASM characteristic towards CON and $CORR$ are shown in Fig. 10. The computational time to display the thermogram results after the final transformation is about 1.906 seconds.

Table 5 describes strong correlations and computational time resulting from second-order statistical features. The lowest computational time is achieved by the pair of the CON and $CORR$ attributes with 1.734 seconds. Whilst the highest computational time is achieved by the pair of the ASM and CON attributes with 5.328 seconds. Overall, the computational time average of all feature pairs is 2.205 seconds. Table 6 presents weak correlations and computational time resulting from second-order statistical features. The only one of weak correlation is obtained by the CON attribute towards the pair of $CORR$ and ENT attributes with computational time in 1.891 seconds.

Table 5. Strong correlations and computational time of second-order statistical feature pair

No	Features Pair	Computational Time (s)
1	ASM and CON	5.328
2	ASM and $CORR$	1.766
3	ASM and END	1.766
4	CON and $CORR$	1.734
5	CON and ENT	1.750
6	$CORR$ and ENT	1.750
7	ASM towards CON and $CORR$	1.906
8	ASM towards CON and ENT	1.906
9	ASM towards $CORR$ and ENT	1.938
	Average	2.205

Table 6. Weak correlations and computational time of second-order statistical characteristic pairs

No.	Features Pair	Computational Time (s)
1.	CON towards $CORR$ and ENN	1.891

Table 7 illustrates the statistical characteristic correlation coefficients. It shows the linear dependence between the two variables x and y, the two variables x and z, as well as the two variables y and z in the second-order statistical characteristic plot. The variables x, y, and z range between -1 and +1 which indicate the degree of linear dependence between the two variables. The results of the first transformation to the thermogram raw data have a correlation coefficient on the variables xt_1, yt_1, and zt_1 indicates that there are no transformations on the x or y variables.

The results of the first transformation to the thermogram raw data have a correlation coefficient on the variables xt_1, yt_1, and zt_1 indicates that there are no transformations on the x or y variables. However, there are variable zt_1 that transformed into the independent variable of the two other variables, which is indicated by the correlation coefficient $zt_1 = 0$.

Similarly, the second transformation results, which has a correlation coefficient on the variables xt_2 , yt_2 , and zt_2 indicates that there are no transformations on the xt_1 or yt_1 variables. However, there are variable zt_2 that transformed into the independent variable of the two other variables, which is indicated by the correlation coefficient $zt_2 = 0$. Weak correlations give incorrect grouping results because it provides redundant information so that there are still some thermograms that will be read as normal thermograms, early cancer thermograms, or advanced cancer thermograms.

Thermogram analysis is used to monitor the examination of breast cancer patients who are undergoing radiation and/or chemotherapy. Monitoring data collection was carried out for three consecutive weeks, which was divided into 3 groups, namely: monitoring1, monitoring2, and monitoring3 of 11 patients who were undergoing radiation and/or chemotherapy examinations.

Using the PCA approach, the monitoring results of the cancer thermogram to the normal thermogram will be obtained. Second-order statistical feature extraction is done by taking the most significant characteristics to distinguish the type of thermogram,

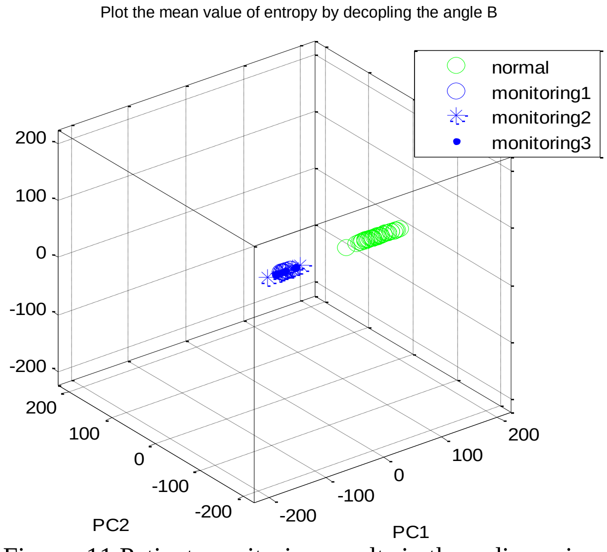


Figure. 11 Patient monitoring results in three dimensions

Table 7. Secondary statistical characteristic correlation coefficients

Feature	x	y	z	xt_1	yt_1	zt_1	xt_2	yt_2	zt_2
$ASM; CON$	-0.329	-0.57	-0.214	-0.329	-0.57	0	-0.329	-0.57	0
$ASM; COR$	-0.014	0.322	0.1379	-0.014	0.322	0	-0.014	0.322	0
$ASM; ENT$	-0.854	-0.9	-0.873	-0.854	-0.9	0	-0.854	-0.9	0
$CON; COR$	-0.337	-0.81	-0.570	-0.337	-0.81	0	-0.337	-0.81	0
$CON; ENT$	0.500	0.82	0.463	0.500	0.82	0	0.500	0.82	0
$COR; ENT$	0.160	-0.63	-0.078	0.160	-0.63	0	0.160	-0.63	0

Received: May 5, 2020. Revised: May 26, 2020.

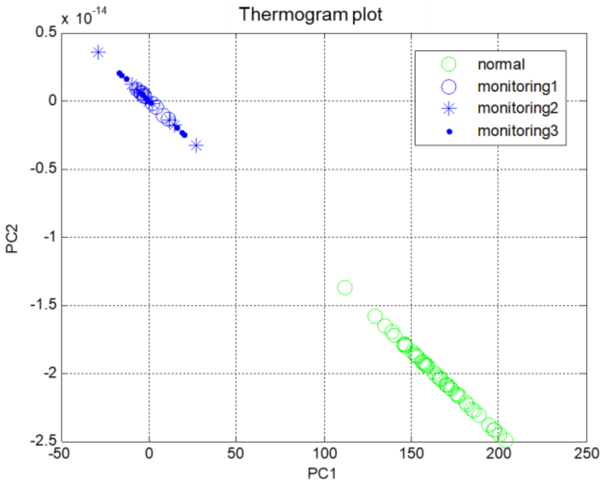


Figure. 12 Thermogram monitoring plot in two dimensions

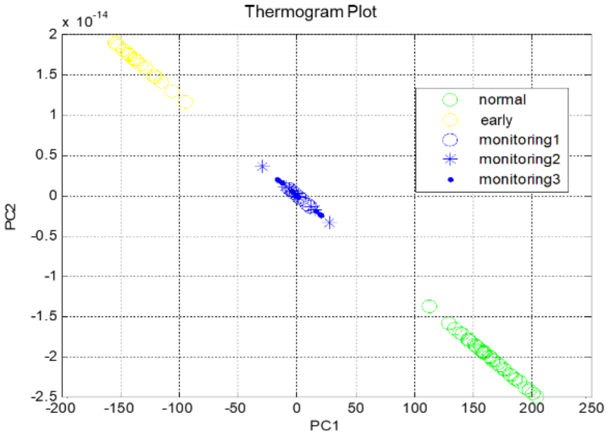


Figure. 13 A plot of thermogram monitoring results for normal and early cancer thermograms

namely the average, and entropy. A three-dimensional plot of the monitoring results of breast cancer patients with a normal thermogram is displayed in Fig. 11. The results of the three-dimensional reduction into two dimensions are the main components of the results of monitoring the cancer thermogram to the normal thermogram describe in Fig. 12.

Using the two most significant features, the results of the normal thermogram plot, the early cancer thermogram against the patient monitoring thermogram are illustrated in Fig. 13. It displays the classification of thermogram types into normal thermogram types, early cancer thermograms, and thermograms of patients undergoing examination (monitored).

The patient thermogram monitoring results for three weeks have not been able to show a difference that is strong enough to separate monitoring1, monitoring2, monitoring3. The PCA technique in this study can be used to separate the normal thermogram

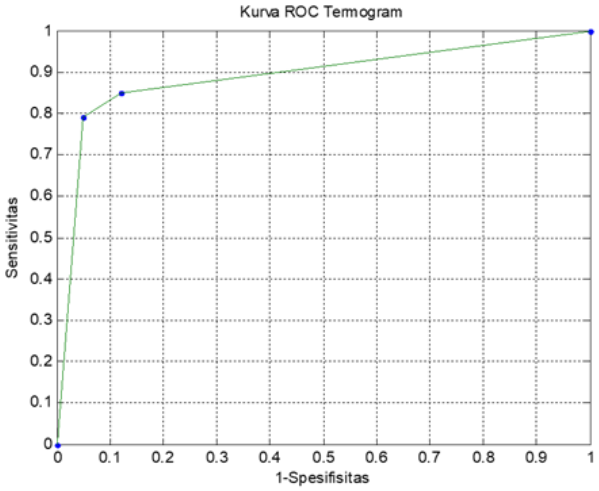


Figure. 14 ROC curve

and the cancer thermogram because it can provide classification results that are far apart from each other [11]. The next step in the analysis results is to calculate the value of sensitivity, specificity, and ROC curve. The sensitivity and specificity parameters in (5) and (6), respectively, can be calculated by:

a. Sensitivity ($Sens$) as a comparison of cancer detected patients with the total number of cancer patients. In this study, the sensitivity value is 92%.
b. Specificity ($Spec$) as a comparison of patients not detected with cancer with the total number of patients not detected with cancer. The specificity value of this study is 95%.

The estimation results of $sens$ and $spec$ can be interpreted as follows:

a. If the test is used for women who do not have breast cancer, then the test will almost certainly be negative. It means that the specificity of 95% is large enough.
b. If the test is used for women suffering from breast cancer, then the chance of being detected is large. It means that the sensitivity of 92% is big enough.

The plot between the sensitivity and 1-specificity values is the ROC curve describes in Fig. 14. It shows that the area under a large ROC curve means that the diagnostic procedure is performed quite well.

For further evaluation, the proposed method of this study is compared to the most related works based on the similar or same platform using PCA that are presented in Table 8. The term of sensitivity and specificity are used for performance evaluation among methods. Based on the average performance in Table 8, it shows that the use of second-order statistical on PCA could yield a promising performance compared to the recent methods. The performance differences between these methods are

Table 8. Comparison of average performance

No.	PCA-based Method	Sensitivity (%)	Specificity (%)
1.	PCA+LDA [3]	83	80
2.	PCA+QDA [3]	83	80
3.	PCA+SVM [15]	99.06	96.56
4.	PCA+SVM [16]	95.97	96.23
5.	PCA+parallel SVM [18]	100	100
6.	PCA+ANN [19]	95	98
7.	PCA+RF [19]	92	97
8.	Fuzzy PCA+SVM [20]	96	98.67
9.	PCA+SVM [20]	85.19	100
10.	PCA+LDA [17]	81	95
11.	PCA+second-order statistics (this study)	92	95

usually influenced by many factors, such as datasets and parameters. Although all of those methods are not re-experimented in this study, however, the proposed method can be used to enrich breast cancer detection with comparable results to the other methods.

6. Conclusions and future works

The PCA with a second-order statistical method through covariance matrix, eigenvalue, and eigenvector has been able to provide a high level of sensitivity and specificity, which are 92% to 95%, in sorting out normal thermograms, early cancers, advanced cancers, and for monitoring the development of therapeutic results in breast cancer patients. Thermogram parameters are extracted from second-order statistical features to represent normal, early and advanced cancers are available. Hence, any other digital thermogram samples will be identified directly using these parameters as the initial composition of the process to shorten computing time. The proposed method that is implemented in the breast thermogram processing also has sufficiently high computing. The computation time of the proposed algorithm is only 2.205 seconds on average.

Several follow-ups that can be taken from this study as future works include:

1. realize the proposed method for real-time and standalone software that can be directly applied to detect breast cancer early and to improve the results of cancer monitoring;
2. using a standard thermal camera specifically for medical applications that have a temperature range of 35º to 40ºC with an accuracy of 0.01ºC.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization and methodology, A. Marjuni, and O. D. Nurhayati; software and validation, O. D. Nurhayati; formal analysis and investigation, A. Marjuni, and O. D. Nurhayati; resources and data curation, O. D. Nurhayati; writing—original draft preparation, A. Marjuni; writing—review and editing, A. Marjuni; visualization, O. D. Nurhayati; supervision, A. Marjuni; project administration, A. Marjuni, and O. D. Nurhayati. All authors have read and approved the final manuscript.

References

[1] WHO, "*Breast Cancer*", Online: Available at https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/, 2018.

[2] A. Omar, A. Bakr, and N. Ibrahim, "Female Medical Students' Awareness, Attitudes, and Knowledge about Early Detection of Breast Cancer in Syrian Private University", *Heliyon*, Vol. 6, Issue. e03819, pp. 1-7, 2020.

[3] F. M. Lyng, D. Traynor, T. N. Q. Nguyen, A. D. Meade, F. Rakib, R. Al-Saady, E. Goormaghtigh, K. Al-Saad, and M.H. Ali, "Discrimination of Breast Cancer from Benign Tumours using Raman Spectroscopy", *PLoS One*. Vol. 14, pp. 1-13, 2019.

[4] WHO, "*Who Position Paper on Mammography Screening*", 2015.

[5] M. A. -Nasser, A. Moreno, and D. Puig, "Breast Cancer Detection in Thermal Infrared Images using Representation Learning and Texture Analysis Methods", *Electron*. Vol. 8, No. 100, pp. 1-18, 2019.

[6] J. Z. -Gomez, N. Zerhouni, Z. Al Masry, C. Devalland, and C. Varnier, "A Survey of Breast Cancer Screening Techniques: Thermography and Electrical Impedance Tomography", *Journal of Medical Engineering and Technology*, Vol. 43, No. 5, pp. 305-322, 2019.

[7] F. AlFayez, M. W. A. E. -Soud, and T. Gaber, "Thermogram Breast Cancer Detection: a Comparative Study of Two Machine Learning Techniques", *Applied Sciences*, Vol. 10, No. 2, pp. 1-20, 2020.

[8] S. V. Francis, M. Sasikala, and S. Saranya, "Detection of Breast Abnormality from Thermograms using Curvelet Transform Based Feature Extraction", *Journal of Medical Systems*, Vol. 38, No. 23, pp. 1-9, 2014.

[9] J. Prakash, “Breast Cancer Detection using Thermogram: Review of Latest Techniques”, *International Journal of Engineering Research & Technology*, Vol. 4. No. 15, pp. 1-12, 2016.

[10] P. Gomathi and V. Jamuna, “Feature Extraction of Thermal Image for Breast Cancer Analysis”, *International Journal of Pure and Applied Mathematics*, Vol. 118, No. 120, pp. 707-712, 2018.

[11] O. D. Nurhayati, T. S. Widodo, A. Susanto, and M. Tjokronagoro, “First Order Statistical Feature for Breast Cancer Detection using Thermal Images”, *World Academy of Science, Engineering and Technology*, Vol. 70, pp. 1040-1043, 2010.

[12] A. Ibrahim, S. Mohammed, and H. A. Ali, “Breast Cancer Detection and Classification using Thermography: A Review”, In: *Proc. of International Conference of Advanced Machine Learning Technologies and Applications*, Cairo, Egypt, pp. 496-505, 2018.

[13] O. D. Nurhayati, A. Susanto, T. S. Widodo, and M. Tjokronagoro, “Principal Component Analysis combined with First Order Statistical Method for Breast Thermal Images Classification”, *International Journal of Computer Science and Technology*, Vol. 2, No. 2, pp. 12-18, 2011.

[14] M. U. Ali, S. Ahmed, J. Ferzund, A. Mehmood, and A. Rehman, “Using PCA and Factor Analysis for Dimensionality Reduction of Bio-informatics Data”, *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 5, pp. 415-426, 2017.

[15] A. Jamal, A. Handayani, A. A. Septiandri, E. Ripmiatin, and Y. Effendi, “Dimensionality Reduction using PCA and K-Means Clustering for Breast Cancer Prediction”, *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, Vol. 9, No. 2, pp. 192-201, 2018.

[16] J. Liu and W. Ma, “An Effective Recognition Method of Breast Cancer Based on PCA and SVM Algorithm”, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNCS 4901, pp. 57-64, 2008.

[17] V. Moisoiu, A. Socaciu, A. Stefancu, S. D. Iancu, I. Boros, C. D. Alecsa, C. Rachieriu, A. R. Chiorean, D. Eniu, N. Leopold, C. Socaciu, and D. T. Eniu, “Breast Cancer Diagnosis by Surface-Enhanced Raman Scattering (SERS) of Urine”, *Applied Sciences*, Vol. 9, No. 4, pp. 1-10, 2019.

[18] F. J. L. -Rosas, J. C. M. -Romo, R. M. -González, H. L. -García, M. A. R. -Díaz, and L. C. R. -Martínez, “PCA and Parallel SVM to Optimize the Diagnostic of Breast Cancer Based on Raman Spectroscopy”, *DYNA New Technologies*, Vol. 5, No. 1, pp. 1-13, 2018.

[19] B. Sahu, S. Mohanty, and S. Rout, “A Hybrid Approach for Breast Cancer Classification and Diagnosis”, *ICST Transactions on Scalable Information Systems*, Vol. 6, No. 20, pp. 1-8, 2018.

[20] M. F. Dzulkalnine, R. Sallehuddin, Y. Yusoff, N. H. M. Radzi, and N. H. Mustaffa, “Fuzzy PCA and Support Vector Machines for Breast Cancer Classification”, *International Journal of Engineering & Technology*, Vol. 7, No. 3.7, pp. 62-64, 2018.

[21] C. Bugli and P. Lambert, “Comparison between Principal Component Analysis and Independent Component Analysis in Electroencephalograms Modelling”, *Biometrical Journal*, Vol. 49, No. 2, pp. 312-327. 2007.

[22] R. Reris and J. P. Brooks, “Principal Component Analysis and Optimization: A Tutorial”, In: *Proc. of INFORMS Computing Society Conference*, Richmond, Virginia, pp. 212-225, 2015.

[23] A. Tharwat, “Principal Component Analysis - a Tutorial”, *International Journal of Applied Pattern Recognition*, Vol. 3, No. 3, pp. 197-240, 2016.

[24] M. Bramer, *Principles of Data Mining: Second Edition*, Springer-Verlag, London, 2013.

[25] W. Hernandez and A. Mendez, *Application of Principal Component Analysis to Image Compression*, In: Statistics - Growing Data Sets and Growing Demand for Statistics, 2018.

[26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Second Edition, John Wiley & Sons, Ltd., 2012.

[27] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Third Edition, Pearson Prentice Hall, 2008.

[28] K. I. Satoto, O. D. Nurhayati, and R. R. Isnanto, “Pattern Recognition to Detect Breast Cancer Thermogram Images Based on Fuzzy Inference System Method”, *International Journal of Computer Science and Technology*, Vol. 2, No. 3, pp. 484-487, 2011.

[29] K. H. -Tilaki, “Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation”, *Caspian Journal of Internal Medicine*, Vol. 4, No. 2, pp. 627-635, 2013.