Multi-View Combination using Mutual Information and 3-D Euclidean Distance for Breast Cancer Classification

Orawan Chunhapran*, Duangjai Noolek*,
Parinda Labcharoenwongs*, and Tongjai Yampaka*

Received: September 1, 2022
Revised: November 18, 2022
Accepted: November 25, 2022

* Corresponding Author: Duangjai Noolek, E-mail: duangjai_noo@rmutto.ac.th

Abstract

The most popular method of early breast cancer detection is mammography, which uses two views: the Medio Lateral Oblique (MLO) and the Cranio Caudal (CC). In practice, experienced radiologists interpret both mammography views in order to categorize them as normal or abnormal. However, human error has been found in classification. This study proposes multi-view combination using mutual information and 3-D Euclidean distance for breast cancer classification. The public dataset Breast Cancer Digital Repository (BCDR) including 600 CC-view and 600 MLO-view was used in this study. Our method divides into five steps. First, pre-training with deep convolutional network was used to extract the significant feature. Second, principal component analysis (PCA) was simultaneously computed the principal components. Third, mutual information (MI) was measured the mutual dependence between components and class label for selecting the best component group. Fourth, 3-dimensional Euclidean distance merging was established to merge both views. Finally, the support vector machine was performed to classify breast lesion in normal, benign or malignant. The model accuracy is 99.33%, and AUC is 0.98. The results demonstrate that the performance of our strategy is more improved when compared with other combination studies.

Keywords: Multi-View Learning, Data Integration, Data Fusion, Breast Cancer Diagnosis.

1. Introduction

Mammography is widely used to detect abnormal breast mass. In practice, the radiologists manually interpret both views simultaneously. However, the performance in the manual analysis is less in a specificity of 91% and a sensitivity of 84% [1]. Single-view and double-view mammographic examination by well-trained radiologists were compared in many studies [2], [3], [4] and reported high detection rate. Klein et al. [5] reviewed some pitfalls in CC-view and MLO-view, then, they suggested that the pitfalls may reduce by finding the image correlation or image integration from both views. According to previous reports [6], it is possible to share information between MLO-view and CC-view for breast cancer classification. The most developments of Computer Aid Diagnosis have widely used machine learning based on Deep Convolutional networks (DCNNs) [7], [8].

Multi-view learning has been introduced to integrate the heterogeneous input view. In medical diagnosis, the decision features are derived from multiple medical evidence and integrated into a final decision. The many simple approaches apply concatenation method to fuse the data, but concatenation features further increase the high-dimensional problem [9]. Single value decomposition (SVD), principal component analysis (PCA), or canonical correlation analysis (CCA) have been used to find a set of new low dimensional space. Therefore, this study proposed multi-view combination using mutual information and 3-D Euclidean distance for breast cancer classification.

* *Department of Computer Science, Faculty of Business Administration, and Information Technology, Rajamangala University of Technology Tawan-ok.*

2. Literature survey

There are many difference approaches of obtaining the multi-view classification and it is related to many research fields. Jouirou et. al, [10] developed a method for extraction and fusion from both CC and MLO views. Their experimental results showed that the fusion of the CC and MLO views might improve the rates of the descriptors evaluated. Different fusion strategies were designed into two schemes [11]. First, Soft Decision Fusion Schemes make use of the posterior probability estimates of the single-source classifiers, while Hard Decision Fusion Schemes performed rules to fuse on the hard decisions. These fusion schemes demonstrate that multi-view strategy performance was improved when compared with single-view systems. The suitable feature fusion was introduced by Sasikala et. al., 2022 [12] using Canonical Correlation Analysis (CCA). According to the observation, decomposition projection schemes such as SVD, PCA, or CCA have been investigated for combining information from CC-view and MLO-views to improve classification performance. However, there are currently two main limitations. The first limitation, deciding the number of components, many studies argued that keeping all components were unnecessary [13]. In practice, only the k-components with high variance scores were used in further analysis. Another selection method is based on the proportion of the total variance explained in 70% to 90% [14]. Graphical approaches [15] suggested the eigenvalues scree plot. The eigenvalues of each component are plotted and applied a straightedge to the bottom portion of the eigenvalues. The values of k are given by the point at above the straight line were chosen [16]. However, the principal eigenvectors were not considered only in component selection but also should be considered the consistent target class label.

Therefore, we proposed the mutual information (MI) measuring the mutual dependence between component and class label. The second limitation, previous combination approaches proposed a concatenated feature vector, but different sources of information usually have correlated and uncorrelated. Multiple views unlike single view, it concerns correlation and ensures their compatible between multiple views. If carried out the correlation across the views, shared representation is well performed in multi-view learning. Consequently, we proposed the nearest merged method using the 3-D Euclidean distance for breast cancer classification. This approach makes certain that only the same characteristics of variance were merged before the classification task.

3. Preliminary Theories

This section discusses the relevant theories for resolving our methodology, including Convolutional Neural Networks, Principal Component Analysis, Mutual Information, and Euclidean distance, and others.

3.1 Convolutional Neural Networks

Convolutional neural networks are a specific subset of artificial neural networks that substitute the mathematical operation convolution for general matrix multiplication in at least one of its layers. They are employed in image processing and recognition. Complicated functions when used deep architectures could be represented in high-level abstractions [17]. However, the depth layers were affected with learning time, then, This problem can be sole using the pre-trained model that reported in large-scale image and video recognition [18-22]. Figure 1 shows the proposed architecture.

3.2 Principal Component Analysis (PCA)

Before starting with PCA, a foundation using in PCA was explained such as covariance, eigenvectors, and eigenvalues.

Covariance: Covariance has been used to measure the correlation and distribution between n variables. The formula for covariance could also be written as:

$$\sum x = \frac{1}{n} X^T X \qquad (1)$$

Eigenvectors and Eigenvalues: Eigenvectors are two multiplications between matrix and vector. For example:

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}, v = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \therefore Av = 4\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \lambda = 4$$

Matrix A can be thought of as a transformation matrix that the eigenvector arises from. Each eigenvector λi has an associated eigenvalue v_i which is the variance of the extracted vector v_i. The eigenvectors and eigenvalues are calculated as:

$$(A - \lambda_i I)v_i = 0, I = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \quad (2)$$

where matrix A is covariance matrix, I is identity matrix. This problem may be solving in linear equations. This process establishes the eigenvectors that involve transforming the data in PCA.

Principal Component Analysis: It is a method to reduce data dimensionality. The original features were eliminated possibly correlated variables and combined them into a smaller number of principal components which project data points into the directions of maximal variance within new space. Suppose that the dataset is represented in a matrix. The dataset matrix $X^i_{n\times q}$ with n points and q features. Each row of X is a data point and each column is a feature. The j point is defined as a q dimension column vector x_j, for j = 1,…,n and the data mean vector is:

$$X^i = \begin{bmatrix}x_1^T\\ x_2^T\\ \vdots\\ x_n^T\end{bmatrix} \quad (3)$$

$$x_j = \begin{bmatrix}x_{j1}\\ x_{j2}\\ \vdots\\ x_{jq}\end{bmatrix} \quad (4)$$

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \quad (5)$$

$$X = \begin{bmatrix}(x_1 - \bar{x})^T\\ (x_2 - \bar{x})^T\\ \vdots\\ (x_n - \bar{x})^T\end{bmatrix} \quad (6)$$

The centered matrix is X having the j^{th} row equal to $(x_j - \bar{x})^T$, The covariance matrix of X is defined as Equations 1, Each column V^i_γ, for $i = 1,\ldots,k$ of the matrix.

$$V^i_\gamma = [v^1_\gamma, \ldots, v^k_\gamma] \quad (7)$$

V^i_γ is an eigenvector of $\sum x$. Each eigenvector V^i_γ have an associated eigenvalue λ^i_γ, which is the variance of the extracted feature C^i_γ. The eigenvectors in V^i_γ are sorted, so that $\lambda_1 > \ldots > \lambda_k$. In PCA, the points are projected in the directions of maximal variances, these directions are the eigenvectors of the covariance matrix that has the greatest eigenvalues. The new data matrix $C^i_{n\times k}$ is defined as:

$$C^i_\gamma = X_\gamma V^i_\gamma \quad (8)$$

This matrix is called components C^i_γ which each row of this matrix is a point and each column an extracted feature.

3.3 Mutual Information

The mutual information *(MI)* for classification was defined as the probability of cross relation between components and opposite class labels, then the *MI* can be maximized the probability of the classification that defined in

$$\boldsymbol{I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)} \quad (9)$$

3.4 Euclidean Distance

The distance between variables could be considered the relationship among them. Euclidean space is represented by Euclidean metric that ordinary straight-line distance between two points. Two points p and q, if $p = (p_1, p_2,..., p_n)$ and $q = (q_1, q_2,..., q_n)$ are two points, distance (*dst*) from p to q is given by the Pythagorean formula.

$$dst = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (10)$$

The nearest distance quantifies the similarity between two objects. This is importance to merge two correlation objects. The nearest operation between p_i and q_i can be defined as:

$$nearest_{dst} = argmin\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (11)$$

4. Research Methodology

The design and development of Multi-view combination in solving breast cancer classification consist of five steps.

a) Input image: the input images were fed to the training step (Fig. 1a).

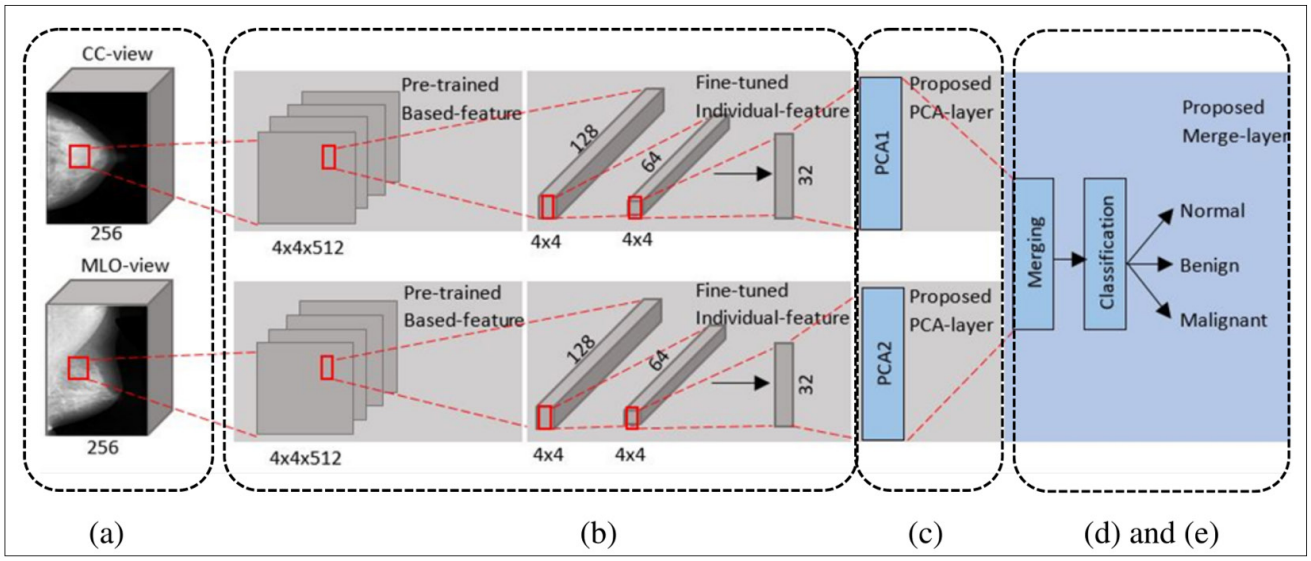


Figure 1. *VGG16 Pre-Trained Layers was First Applied with each Input Images Called based Features. The Output Features from these Layers Called Individual Features. Then, Output Feature Feed to PCA Layer Follow by Classification Layer.*

b) CNN: the pre-trained model was trained the training dataset (fig 1b) followed by CNN top layer for learning individual feature.

c) PCA: the PCA was used for dimensionality reduction by projecting each data point onto only the first few principal components (fig 1c) followed by feature selection using mutual information technique for selecting the relevant features.

d) Information merging: the Euclidean distance was used to combine the principle component instead of concatenation method. (fig 1d)

e) Classification: Support Vector Machine was classified the merging feature as normal, benign, or malignant.

4.1 Materials

Digital Database for Screening Mammography (DDSM) [23] (600 CC-view and 600 MLO-view) was used in this study. There are divided into three class normal, benign, or malignant. The dataset was slit 70% for training and 30% for testing. Figure 2 shows the sample MLO view and CC view breast images.

4.2 Feature Extraction

VGG16 pre-trained layers was first applied with each input images called based features. Because of the heterogeneous dataset, the top-model layers were individually designed and

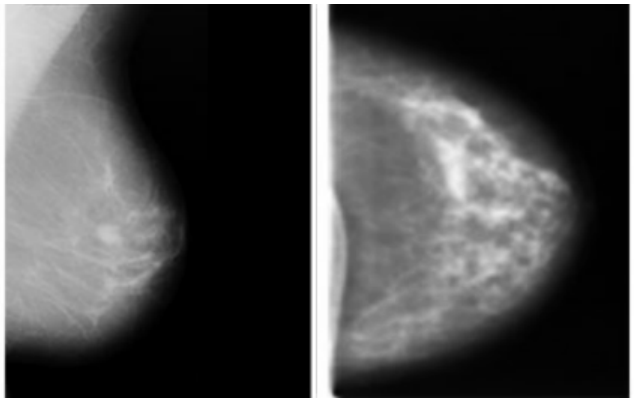


Figure 2. *Sample Mammographic Image in MLO View and CC View.*

fine-tuned to apply with based features. The output features from these layers called individual features that were ready to use the next step.

4.3 Principal Component Analysis (PCA)

PCA process can be done follow:

a) Form two datasets into the matrix using Eq. (3). Let $X^1_{n\times q}$ be the first dataset, while $X^2_{n\times q}$ be the second dataset.

b) Separately compute mean vector along each matrix using Eq. (5).

c) Centered the data by subtract each column vectors $X^1_{n\times q}$ and $X^2_{n\times q}$ by their mean vectors using Eq. (6).

d) Separately compute covariance matrix using Eq. (3).

e) Separately compute eigenvector using Eq. (7), V_{x1}

from $X^1_{n\times q}$ and V_{x2} from $X^2_{n\times q}$ that correspond with their eigenvalue λ_{x1} and λ_{x2}.

f) Separately establish C^i_y using Eq. (8), the first dataset components defined as C^i_x, while the second components defined as C^i_y. From the Eq. (8), it should be defined in the linear combination of the original feature is written as:

$$C^i_x = \alpha^i_{11}x_1 + \alpha^i_{12}x_2 + \cdots + \alpha^i_{1p}x_p$$
$$C^i_y = \beta^i_{11}y_1 + \beta^i_{12}y_2 + \cdots + \beta^i_{1q}y_q$$

The weights α and β are obtained from the eigenvector $V^i_{\times 1}$ and $V^i_{\times 2}$ associated eigenvalue $\lambda^1_y \lambda^1_y$, which is the variance of the extracted feature.

The eigenvectors in V^i_γ are sorted, so that $\lambda 1 > \ldots > \lambda_k$. The first component is the largest variance of X_1 and X_2, while the second component is lower variance than the first. The remaining components were defined with $C^1_\gamma > C^2_\gamma > C^i_\gamma$. The variance scores are defined in eigenvalues, Let $\lambda^1_\gamma \geq \lambda^2_\gamma \geq \cdots \geq \lambda^i_\gamma \geq 0$. The obtained principal component based on the number of features. Let $\lambda^1_\gamma \geq \lambda^2_\gamma \geq \cdots \geq \lambda^p_\gamma$ for the first dataset as similar as the second dataset $\lambda^1_\gamma \geq \lambda^2_\gamma \geq \cdots \geq \lambda^q_\gamma$

In practice, only the k components with high variance scores were used in further analysis. Previous works as followed by David A. Ratkowsky [24] suggested the *k-components* that larger than average eigenvalues. Another selection method is based on the proportion of the total variance explained in 70% to 90% [25]. Graphical approaches [26], [27] suggested the scree plot which plotted and applied a straightedge to the bottom portion of the eigenvalues, then, the values of k are given by the point at above the straight line were chosen. However, the component of the highest eigenvalue was less suitable for classification than another component [28]. Therefore, we proposed a method to select components by mutual information schemes.

4.4 Estimating Mutual Information of Principal Components

The mutual information of two datasets can be calculated from equation (9).

$$CMI(\gamma; l_\gamma) = \sum_{l_\gamma \in L} \sum_{\gamma \in C^i_x, C^i_y} p(\gamma, l_\gamma) \log\left(\frac{p(\gamma, l_\gamma)}{p(\gamma)p(l_\gamma)}\right) \geq mean_{CMI_\gamma}$$

Where γ be C^i_γ, l_γ be target class of $X^1_{n\times q}$ and $X^2_{n\times p}$. For any particular value of component, a low probability means that outcome is less likely to occur, and these variables should not be appearing, while a high probability variable should be appearing for classification, then, the components which have *MI* scores over mean were included in group.

4.5 Merging using 3-Dimensional Euclidean Distance

For the process followed by principal component analysis and selected component method, the merging process was performed using 3-dimention Euclidean distance between the significant points of components. The 3-Dimensional Euclidean distance was calculated between any three components points in space corresponds to the length of a straight line drawn between them. Consider a collection of 3 points $\{C^1_x, C^2_x, C^1_3\}$ are the first 3 components of the first dataset and $\{C^1_y, C^2_y, C^1_y\}$are the first 3 components of the second dataset. Followed by Equation (10):

$$dst_i = \sqrt{(C^1_x - C^1_y)^2 + (C^2_x - C^2_y)^2 + (C^3_x - C^3_y)^2}$$

$$nearest_{dst} = argmin\sqrt{\sum_{i=1}^{n}(C^i_x - C^i_y)^2}$$

4.6 Classification

The classification stage was performed to classify tumour lesion using Support Vector Machine to categorize the breast mass either as normal, benign, or malignant. The k-fold cross validation procedure was used to obtain the performance evaluated by Confusion matrices including sensitivity (true positive rate), specificity (true negative rate), over all accuracy, and ROC.

5. Experiment Results

5.1 Result in Principal Component Analysis

The PCA aimed to extract the most important features. Table 1 shows the total variance of CC-view and MLO-view. Both views show five principle components with eigenvalues greater than 1.0. The percentage of cumulative shows 90% explains variance. The scree plot as show in figure 3 represented

Table 1. *Total of Variance Explain of CC-View and MLO-View Features.*

Total Variance Explain						
Component	CC-view			MLO-view		
	Total	%Variance	% Cumulative	Total	%Variance	% Cumulative
1	17.779	48.052	48.052	19.561	44.456	44.456
2	12.239	33.078	81.129	16.907	38.425	82.880
3	1.631	4.409	85.538	1.839	4.180	87.060
4	1.213	3.278	88.815	1.352	3.073	90.133
5	1.079	2.915	91.731	1.227	2.788	92.921

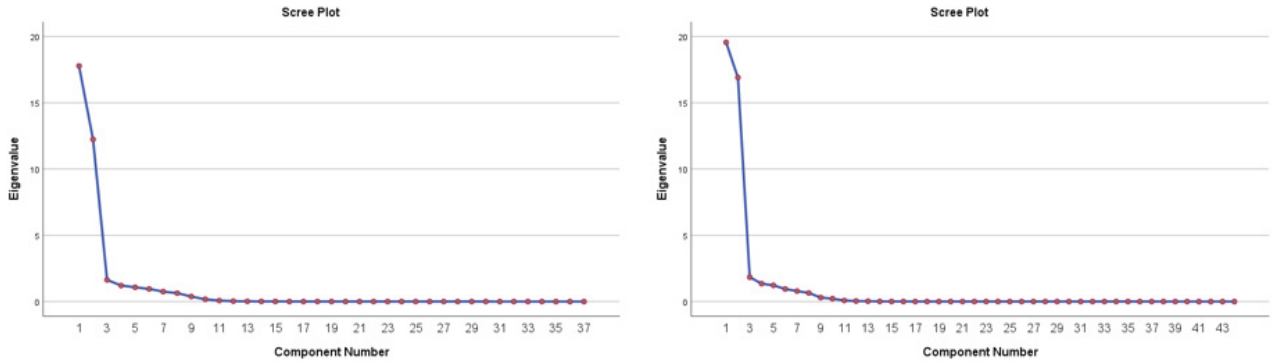
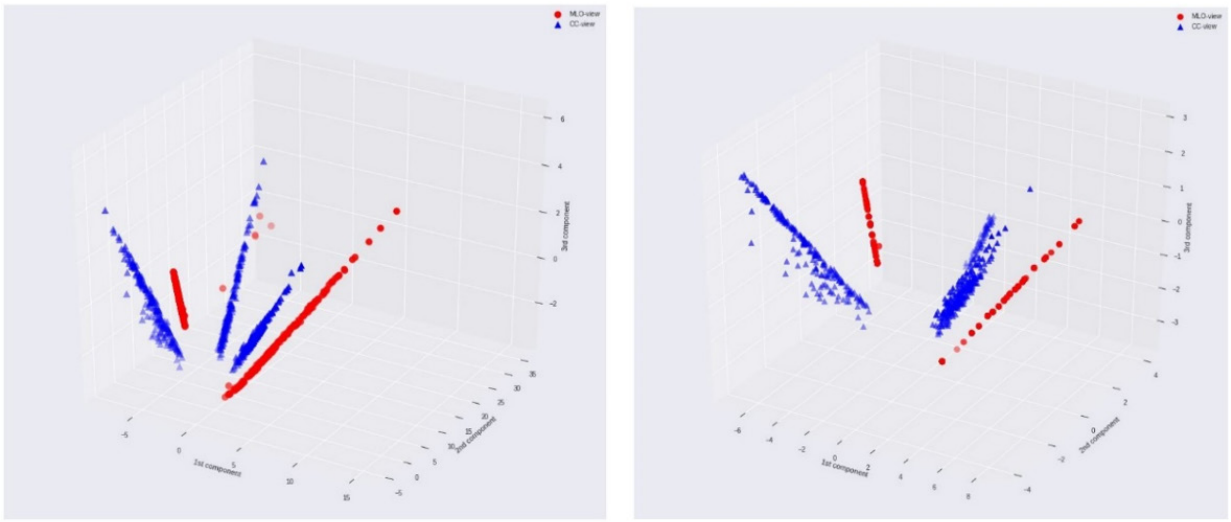


Figure 3. *Scree Plot of CC-View and MLO-View.*



(a) (b)

Figure 4. *The Merging Comparison of CC-View and MLO-View using all Compnents (a) and Nearest Component using 3-D Euclidean Distance (b).*

Table 2. *The Performance Comparison with Other Researches.*

Paper	System	Sensitivity	Specificity	Accuracy	AUC
Bekker et. al (2016) [29]	Model based	78.80	78.7	78.7	0.89
Carneiro et. al. (2017) [30]	Model based	-	-	-	0.96
Shen et. al. (2019) [31]	Model based	86.10	80.10	-	0.91
Sasikala et. al. (2018) [32]	PCA	93.00	91.18	92.50	-
Sasikala et. al. (2018) [33]	CCA	96.60	95.60	96.10	-
Proposed	**PCA+MI+3D**	**99.00**	**99.50**	**99.33**	**0.98**

the variance scores. The components that have high variance (more than 1) are popularly selected for the classification process. When the feature selection is affected with the performance of classification process, the consistent between feature and target class label should be considered. Then, the mutual information scores over mean of each component were selected. Figure 4 shows MI scores of two views.

5.2 Merging of Information

The 3-Dimensional Euclidean distance was calculated between any three components points in space corresponds to the length of a straight line drawn between them. The results showed that only the nearest pair were merged and used in the classification process. Figure 5 shows the component distribution in 3-dimensional Euclidean space. All components merging (Figure 4a) and nearest component merging (Figure 4b) was compared. The comparison result show that the nearest component merging of CC-view and MLO-view is closer than all component merging.

5.3 Classification Performance

Mammography dataset was classified into three classes (Normal, Benign, of Malignant). The SVM was used to classify this merged dataset. Sensitivity, specificity, and overall performance are 99.00, 99.50, and 99.33 respectively (Figure 5). The AUC is 0.98 (Figure 6).

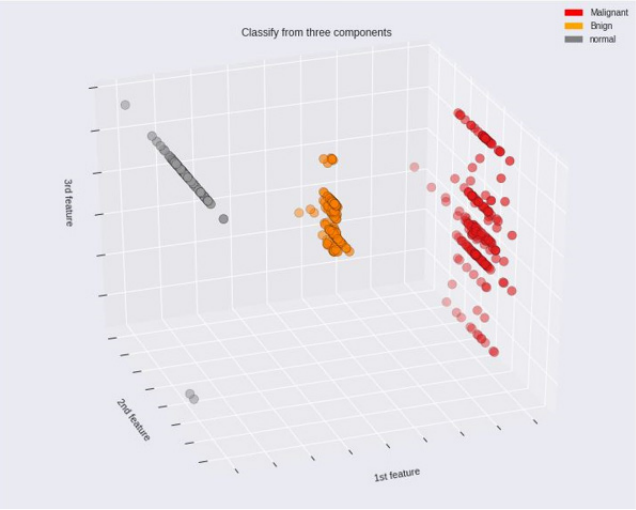


Figure 5. *The Classification Performance.*

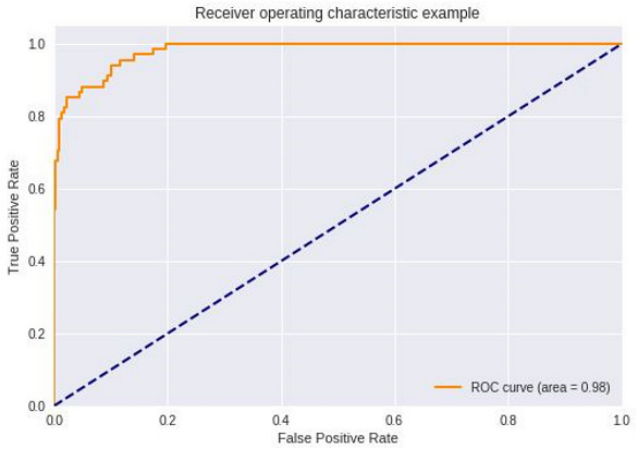


Figure 6. *ROC Curve of Combination Dataset.*

Our approach was compared with the previous works (Table 2). Many studies show the high performance classification when using multiple views of breast mammogram. In addition, this study found that using the feature selection method is outperform using all features. A few recent studies proposed matrices decomposition techniques such as SVD, PCA, or CCA followed by the classification stage and showed the good performance. Generally, the values of k-components are chosen by the high eigenvalues. To extend from the previous studies, the mutual information score between component and class label were included in group. In addition, the most studies applied concatenation method to merge the principal components. Nevertheless, not at all components were concatenated because datasets were statistically different and inequality dimensions. Therefore, the merging using 3-dimention Euclidean distance between the significant points of components could be achieved in better performance compared to the existing method.

6. Discussion and Conclusion

The high data dimension that affected with learning performance is widely reduce using Principal Component Analysis (PCA). The objective of this work not only reduced data dimension but also considered regarding the consistency between components and class labels. Therefore, the supervised PCA have been proposed including mutual information to extend the standard PCA. We also found that the components including with MI could be more sustainable against the class labels. These results are consistent with previous research [34]. They defined class representatives and computed PCA for these points. Other studies, the posterior probability was introduced [35] selected the same features as PCA but selected the ones that minimized the Bayes error rate, while standard PCA selected the features with maximal variance. These studies suggested that components of maximal variance might not be the single way to separate data from different classes.

The effective components were selected from each view followed by merging method followed by principal component analysis. Then, the concatenation method was used to merge among two component groups. In contrast with these methods [32], [33] concatenation component method has not satisfied with statistically different and inequality dimensions of datasets. Consequently, our work proposed the nearest merged method using the 3-dimension Euclidean distance instead general concatenation method. It significantly increased the diversity of records. We also found that the abundant of training data may improve the model accuracy.

The present study was designed to combine the feature from CC-view and MLO-view. In agreement with [36] demonstrated that learning from multiple view would be better than single view. This is supported by [5] reviewed some common pitfalls in breast image and suggested to explore correlation of image or integrate of double reading. This study presented the methodology of multi-view learning for breast cancer diagnosis. Our proposed achieved in better performance compared to the existing method. Although early cancer diagnosis is the key to improve the patient quality of life, the false positive and false negative are appearing. Therefore, accurate and reliable tool will become developed to help the clinician decision. Our experiment indicated that false positive and false negative tended to reduce, furthermore, overall accuracy is better when compared with other strategies. For future works, the proposed method can be extended to the problem with other datasets such as patient demographics, health history record, ultrasound image, or pathological image. Another research investigation is to test the method with not only cancer diagnosis but also cancer prognosis should be explored.

7. References

[1] M. Giger, M. Inciardi, A. Edwards, J. Papaioannou, K Drukker, Y. Jiang, R. Brem, and J. Brown, Automated Breast Ultrasound in Breast Cancer Screening of Women With Dense Breasts: Reader Study of Mammography-Negative and Mammography-

Positive Cancers. *American Journal of Roentgenology,* Vol. 206, No. 6, pp.1341-1350, 2016.

[2] L. Bassett, J. Diamond, R. Gold, and R. McLelland, "Survey of mammography practices", *American Journal of Roentgenology,* Vol. 149, No. 6, pp. 1149-1152, 1987. Available: 10.2214/ajr.149.6.1149.

[3] R. Blanks, M. Wallis, and S. Moss, "A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK National Health Service breast screening programme". *Journal of Medical Screening,* Vol. 5, No. 4, pp. 195-201, 1998. Available: 10.1136/jms.5.4.195.

[4] M. Michell, R. Wasan, A. Iqbal, C. Peacock, D. Evans, and J. Morel, "Two-view 2D digital mammography versus one-view digital breast tomosynthesis". *Breast Cancer Research,* Vol. 12, No. 3, 2010. Available: 10.1186/bcr2656.

[5] S. Roberts-Klein, E. Iuanow, and P. Slanetz, "Avoiding Pitfalls in Mammographic Interpretation". *Canadian Association of Radiologists Journal,* Vol. 62, No. 1, pp. 50-59, 2011. Available: 10.1016/j.carj.2010.07.004.

[6] M. Popli, R. Teotia, M. Narang, and H. Krishna, "Breast Positioning during Mammography: Mistakes to be Avoided". *Breast Cancer: Basic and Clinical Research,* Vol. 8, p. BCBCR.S17617, 2014. Available: 10.4137/bcbcr.s17617.

[7] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge". *International Journal of Computer Vision,* Vol. 115, No. 3, pp. 211-252, 2015. Available: 10.1007/s11263-015-0816-y.

[8] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?". *IEEE Transactions on Medical Imaging,* Vol. 35, No. 5, pp. 1299-1312, 2016. Available: 10.1109/tmi.2016.2535302.

[9] M. Salvi, and F. Molinari, "Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images". *BioMedical Engineering OnLine,* Vol. 17, No. 1, 2018. Available: 10.1186/s12938-018-0518-0.

[10] A. Jouirou, A. Baâzaoui, and W. Barhoumi, "Multi-view information fusion in mammograms: A comprehensive overview". *Information Fusion,* Vol. 52, pp. 308-321, 2019. Available: 10.1016/j.inffus.2019.05.001.

[11] R. Lavanya, and N. Nagarajan, "Information Fusion in CAD Systems for Breast Cancer Diagnosis Using Mammography and Ultrasound Imaging: A Survey". *Journal of Artificial Intelligence,* Vol. 7, No. 3, pp. 113-122, 2014. Available: 10.3923/jai.2014.113.122.

[12] Y. Mao, H. Lim, M. Ni, W. Yan, D. Wong, and J. Cheung, "Breast Tumour Classification Using Ultrasound Elastography with Machine Learning: A Systematic Scoping Review". *Cancers,* Vol. 14, No. 2, pp. 367, 2022. Available: 10.3390/cancers14020367.

[13] S. Huang, Y. Yang, X. Jin, Y. Zhang, Q. Jiang, and S. Yao, "Multi-Sensor Image Fusion Using Optimized Support Vector Machine and Multiscale Weighted Principal Component Analysis". *Electronics,* Vol. 9, No. 9, pp. 1531, 2020. Available: 10.3390/electronics9091531.

[14] I. Jolliffe, and J. Cadima, "Principal component analysis: a review and recent developments". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* Vol. 374, No. 2065, p. 20150202, 2016. Available: 10.1098/rsta.2015.0202.

[15] R. Ledesma, P. Valero-Mora, and G. Macbeth, "The Scree Test and the Number of Factors: a Dynamic Graphics Approach". *The Spanish Journal of Psychology,* Vol. 18, 2015. Available: 10.1017/sjp.2015.13.

[16] J. Trewn, *"Multivariate statistical methods in quality management," 5.2: Principal Component Analysis Based on Covariance Matrices | Engineering360.* Available Online at https://www.globalspec.com/reference/71521/203279/5-2-principal-component-analysis-based-on-covariance-matrices, accessed on 26 March 2022.

[17] Y. Le Cun, Y. Bengio, and G. Hinton, "Deep Learning". *Nature,* Vol. 521, pp. 436-44, 2015. 10.1038/nature14539.

[18] G. Litjens et al., "A survey on deep learning in medical image analysis". *Medical Image Analysis,* Vol. 42, pp. 60-88, 2017. Available: 10.1016/j.media.2017.07.005.

[19] G. de la Cruz, Y. Du, and M. Taylor, "Pre-training with non-expert human demonstration for deep reinforcement learning", *The Knowledge Engineering Review,* Vol. 34, 2019. Available: 10.1017/s0269888919000055.

[20] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks". *Communications of the ACM,* Vol. 60, No. 6, pp. 84-90, 2017. Available: 10.1145/3065386.

[21] P. Sermanet, D. Eigen, X. Zhang, et al., "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks". *Computer Vision and Pattern Recognition,* pp. 1-16, 2014.

[22] USF Digital Mammography Home Page, *DDSM: Digital Database for Screening Mammography.* Available Online at http://www.eng.usf.edu/cvprg/Mammography/Database.html, accessed on 26 March 2022.

[23] G. Lee et al., "Supervised Multi-View Canonical Correlation Analysis (sMVCCA): Integrating Histologic and Proteomic Features for Predicting Recurrent Prostate Cancer". *IEEE Transactions on Medical Imaging,* Vol. 34, No. 1, pp. 284-297, 2015. Available: 10.1109/tmi.2014.2355175.

[24] D. Ratkowsky, "Choosing the number of principal coordinates when using CAP, the canonical analysis of principal coordinates". *Austral Ecology,* Vol. 41, No. 7, pp. 842-851, 2016. Available: 10.1111/aec.12378.

[25] V. Gray. *Principal Component Analysis.* Hauppauge: Nova Science Publishers, Inc., 2017.

[26] R. Canuto, S. Camey, D. Gigante, A. Menezes, and M. Olinto, "Focused Principal Component Analysis: a graphical method for exploring dietary patterns", *Cadernos de Saúde Pública,* Vol. 26, No. 11, pp. 2149-2156, 2010. Available: 10.1590/s0102-311x2010001100016.

[27] R. Abdesselam, "A Topological Approach of Principal Component Analysis". *International Journal of Data Science and Analysis,* Vol. 7, No. 2, pp. 20, 2021. Available: 10.11648/j.ijdsa.20210702.11.

[28] B. ALKAN, "Robust Principal Component Analysis Based On Modified Minimum Covariance Determinant In The Presence Of Outliers". *Alphanumeric Journal,* Vol. 4, No. 2, 2016. Available: 10.17093/aj.2016.4.2.5000189525.

[29] A. Becker, M. Marcon, S. Ghafoor, M. Wurnig, T. Frauenfelder, and A. Boss, "Deep Learning in Mammography", *Investigative Radiology,* Vol. 52, No. 7, pp. 434-440, 2017. Available: 10.1097/rli.0000000000000358.

[30] G. Carneiro, J. Nascimento, and A. Bradley, "Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning". *IEEE Transactions on Medical Imaging,* Vol. 36, No. 11, pp. 2355-2365, 2017. Available: 10.1109/tmi.2017.2751523.

[31] L. Shen, L. Margolies, J. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography". *Scientific Reports,* Vol. 9, No. 1, 2019. Available:

10.1038/s41598-019-48995-4.

[32] S. Sasikala, and M. Ezhilarasi, "Fusion of k-Gabor features from medio-lateral-oblique and craniocaudal view mammograms for improved breast cancer diagnosis". *Journal of Cancer Research and Therapeutics,* Vol. 14, No. 5, pp. 1036, 2018. Available: 10.4103/jcrt.jcrt_1352_16.

[33] S. Sasikala, and M. Ezhilarasi, "Comparative Analysis of Serial and Parallel Fusion on Texture Features for Improved Breast Cancer Diagnosis". *Current Medical Imaging Reviews,* Vol. 14, No. 6, pp. 957-968, 2018. Available: 10.2174/1573405613666170926164625.

[34] A. Bommert, T. Welchowski, M. Schmid and J. Rahnenführer, "Benchmark of filter methods for feature selection in high-dimensional gene expression survival data", *Briefings in Bioinformatics,* Vol. 23, No. 1, 2021. Available: 10.1093/bib/bbab354.

[35] A. Jouirou, A. Baâzaoui, and W. Barhoumi, "Multi-view information fusion in mammograms: A comprehensive overview". *Information Fusion,* Vol. 52, pp. 308-321, 2019. Available: 10.1016/j.inffus.2019.05.001.

[36] N. Derbel, H. Tmar, and A. Mahfoudhi, "A multi-view deep convolutional neural network for reduction of false positive findings in breast cancer screening," *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2020.