

RESEARCH ARTICLE

OPEN ACCESS

Manuscript received October 20, 2022; revised November 10, 2022; accepted November 20, 2022; date of publication December 20, 2022

Digital Object Identifier (DOI): <https://doi.org/10.35882/teknokes.v15i4.458>

Copyright © 2022 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/))

How to cite Ega Elfira, Wahmisari Priharti, and Dien Rahmawati "Comparison of GLCM and First Order Feature Extraction Methods for Classification of Mammogram Images", Jurnal Teknokes, vol. 15, no. 4, pp. 197–205, December. 2022.

Comparison of GLCM and First Order Feature Extraction Methods for Classification of Mammogram Images

Ega Elfira , Wahmisari Priharti , and Dien Rahmawati 

School of Electrical Engineering, Telkom University

Corresponding author: Ega Elfira (e-mail: egaelfira@student.telkomuniversity.ac.id)

ABSTRACT Breast cancer is one of the main causes of death in women and ranks first in cancer cases in Indonesia. Therefore, an early detection and prevention of breast cancer is necessary, one of which is through mammography procedures. A machine learning classifier such as Support Vector Machines (SVM) could be used as an aid to the doctors and radiologist in diagnosing breast cancer from the mammogram images. The aim of this paper is to compare two feature extraction methods used in SVM, namely the Gray Level Co-Occurrence Matrix (GLCM) and first order with two kernels for each method, namely Gaussian and Polynomial. Classification using SVM method is carried out by testing several parameters such as the value of C, gamma, degree and varying the pixel spacing values in GLCM, which usually in previous studies only used the default pixel spacing. The dataset consists of 500 mammogram images containing 250 benign and malignant images, respectively. This study is expected to find out the best method with the highest accuracy between these two texture feature extractions and able to distinguish between benign and malignant classes correctly. The result achieved that Gray Level Co-Occurrence Matrix (GLCM) feature extraction method with both Gaussian and Polynomial kernel yields the best performance with an accuracy of 89%.

INDEX TERMS Breast Cancer, Mammogram Image, Machine Learning, Support Vector Machines, Gray Level Co-Occurrence Matrix, First Order.

I. INTRODUCTION

Breast cancer is one of the leading causes of death in women. Based on data from the Global Cancer Observatory 2018 revealed that breast cancer cases were 58,255 cases or 16.7% of the 348,809 total cancer cases, making breast cancer the most common cancer in Indonesia [1]. In 2020, breast cancer cases increased by 65,858 new cases with 22,430 deaths [2]. Seeing the number of cases of breast cancer, it is necessary to prevent it by conducting an early examination so as to reduce the number of cases of breast cancer patients. There are supporting examinations that support the detection of breast cancer, one of which is using the mammography method.

Mammography is the process of scanning the compressed breast using low-dose x-rays to view breast tissue or glands [3]. Currently, mammography screening is one of breast

cancer screening method that has proven to be the most effective [4]. Breast screening using the mammography method aims to detect abnormalities in the breast that cannot be touched so that it can anticipate the continuous growth pattern of these abnormalities [5]. The development of technology today has many methods that can be used to assist medical personnel in the detection of breast cancer. There are several methods that have been implemented, for an explanation can be seen in the Table 1. With several methodologies that have been implemented for the identification of breast cancer, different accuracy results are obtained. The first study mentioned in the Table 1, tested several parameters in the ANN method, but only used the default parameters in GLCM and so did the second study [6][7]. Previous studies regarding SVM mentioned in the table obtained the highest accuracy. This studies used wavelet feature and hough transform methods. Study using the

wavelet features tested several statistical intensities, including shape of segmented which has big role for proper diagnosis [9]. While study using hough transform uses additional features in preprocessing, namely the gradient based threshold method to remove unwanted label so it can easily suitable for classification stage [10]. However, from the two studies related to SVM, both of them did not explore the parameters of SVM that could affect the results obtained. Parameters in SVM can influence the decision of the best dividing line or hyperplane to separate between the two classes. Therefore, this study will use the method that obtained the highest accuracy in previous studies, namely SVM methods using different feature extraction and testing several parameters such as the value of C, gamma and degree in SVM.

TABLE 1
Previous Research

Method	Application	Accuracy	Ref.
Feed-forward Backpropagation and Cascade-forward backpropagation neural network with GLCM features	Compare between Feed-forward Backpropagation and Cascade-forward backpropagation NN to differentiate between benign and malignant classes	Feed – Forward Backpropagation NN = 87.5%. Cascade-forward backpropagation NN = 67.8%	[6]
Convolutional Neural Network	This study is suitable approach to enhance the small features like micro-calcification and low contrast features such as masses to differentiate between normal and abnormal classes	65%.	[7]
K-Nearest Neighbour (KNN) Classifier with First Order Statistics, Co-Occurrence Matrix Features, Run Length Matrices Features and Laws' Texture Energy Measures.	Classification based on microclassification (MCs) to distinguish benign and malignant classes	FOS = 79% GLCM = 82% GLRLMs = 62% LTEM = 89%	[8]

Support Vector Machines using feature extraction, include wavelet features, statistical intensity-based features and shape features	Breast mass classification for normal - abnormal and benign – malignan classes of previously segmented mass candidates.	92.3%	[9]
Support Vector Machines and Hough transform	To differentiate between normal and abnormal classes with Hough transform.	94%	[10]

The purpose of this study is to compare two types of texture feature extraction methods, namely the GLCM method and the first order using the SVM classification method in order to obtain a system with the best accuracy that can distinguish between the two benign and malignant classes. This study will go through several stages of image processing, including pre-processing, feature extraction using the Gray Level Co-Occurrence Matrix and first order method, and at the final stage the image will be classified according to benign or malignant class. The results of the system design will be displayed in the GUI in matlab. The results of this study are expected to be able to identify breast cancer effectively and accurately by providing an accuracy value above 80%. The system designed is as a tool to support the identification of breast cancer. The results of the classification system are not necessarily correct and must be consulted with the relevant doctor.

II. METHODOLOGY

Breast cancer classification system consists of four main blocks of system design diagrams, namely mammogram image input, pre-processing, texture feature extraction, and classification explain **FIGURE 1**.

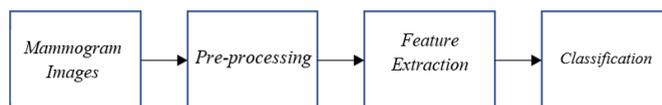


FIGURE 1. Block Diagram

In general, system design has several stages, namely mammogram image input, pre-processing stage to shorten the time in the computation process and make it easier for the next stage. The feature extraction stage aims to retrieve the characteristics of each image using the Gray Level Co-Occurrence Matrix (GLCM) method and order one so that the results obtained can be used in the classification process. The classification process uses the Support Vector Machine (SVM) method.

A. MATERIALS

The dataset used in this study is mammogram image from The Digital Database for Screening Mammography (DDSM). The image used is the image of a woman's breast taken from the left and right breasts. The image used is a grayscale image stored in 8 bit format. The size of each image is 227 x 227 pixels [11]. The image used has 2 classes, namely benign and malignant. The total dataset used is 500 images. There are 400 training images consisting of 200 benign images and 200 malignant images. There are 100 test images consisting of 50 benign images and 50 malignant images.

B. FEATURE EXTRACTION

Features extraction is the process of taking characteristics of an object which will be used as a differentiator from other objects. This characteristic will be used as a distinguishing parameter to describe an object. Parameter values will be used as input in the classification process [12]. Texture feature extraction is taking the characteristics of an object based on information in the form of the surface structure of an image [13]. Texture is the difference between variations and surfaces in an image [14], so that it is able to distinguish between benign and malignant patterns. The texture features that will be used are the first order feature extraction method and the Gray Level Co-Occurrence Matrix (GLCM) method.

1) GLCM METHOD

GLCM is a texture feature extraction method that retrieves information from an image so that it can be used at a later stage. This GLCM is a co-occurrence matrix whose elements describe the number of occurrences of pixels that have a certain gray level value [15]. This GLCM method uses second order calculations based on neighboring calculations between pixels or matrices [16]. Co-occurrence has the meaning as a concurrent event, which means the number of occurrences at one pixel level adjacent to another pixel value based on the spatial distance (d) and the orientation of an angle (θ) [17].

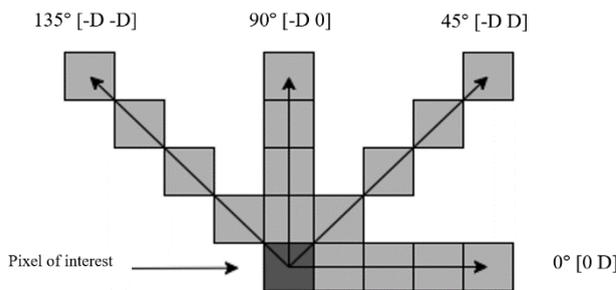


FIGURE 2. The Orientation Direction of GLCM

Spatial distance in this image is denoted in pixels and for angular orientation is denoted in degrees. Spatial relationship is defined as the pixel of interest and the pixel next to it which direction is determined according to specified angle [18]. Basically, GLCM uses four angular orientations, namely 0°, 45°, 90°, and 135° [15]. **FIGURE 2** Shows the orientation direction of the 0° angle represents the reference to the

positive x-axis or horizontal direction. The orientation directions of the 45° and 135° angles represent the reference on the diagonal axis. The orientation direction of the 90° angle represents the reference in the vertical direction. There are four feature parameters used to measure the texture value in this method, there are contrast, correlation, energy and homogeneity.

2) FIRST ORDER

The first order method is one of the methods of taking features from an image based on the histogram characteristics of the image [19]. In contrast to the GLCM method, this first order is based on statistical calculations derived from the original image pixel value and does not pay attention to neighboring pixels [20]. The histogram indicates the possible occurrence of the pixel gray level value in the image. The values obtained in the resulting histogram can then be calculated several feature parameters. There are four feature parameters used to measure the texture value in this method, there are mean, entropy, variance and skewness [21].

C. CLASSIFICATION

This paper presents breast cancer identification from mammogram images using Support Vector Machine. Support Vector Machine (SVM) is a supervised learning method used for classification and regression. SVM aims to find the best dividing line or hyperplane margin that separates the two classes in the input space [22]. SVM works by maximizing the margin between the hyperplane and training data [23]. The basic concept of SVM is to find this hyperplane based on support vectors and margins [24]. Support vectors are all data vectors that are closest to the hyperplane, while the margin is the width of the separating hyperplane [25].

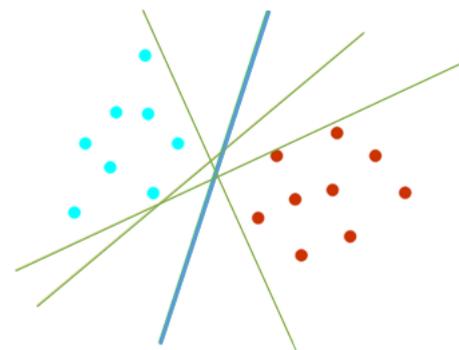


FIGURE 3. Optimal separating hyperplane

In **FIGURE 3** there is a thick blue line in the middle. The blue line shows the best hyperplane between the two classes. For red circles and blue circles that hit the dotted line, it is called a support vector. It shows some data that are members of two different classes. Two classes separated by a hyperplane so as to obtain the following equation:

$$x_i w + b \geq +1 \quad \text{for } y_i = +1 \quad (1)$$

$$x_i w + b \leq -1 \quad \text{for } y_i = -1 \quad (2)$$

Where x is a vector of the dataset that mapped to a high dimensional space. w and b are parameters of the hyperplane that will be estimated by SVM. The basic principle of SVM is linear which is then extended so that it can work on non-linear problems using the kernel trick concept [26]. The kernel function aims to map the initial dimensions of the data set to the new dimensions [27]. In this study, there are two types of kernel functions including Gaussian and Polynomial.

Gaussian kernel is a kernel that has good performance with certain parameters. The Gaussian kernel is defined in the domain of infinite cardinality without any limitation on the number of training samples so as to produce a feature space with infinite dimensions [28]. The results of the training of this kernel have a small error value compared to other types of kernels [29]. In performing the analysis with this kernel, the cost and gamma parameters will be optimized. Cost (c) is used to avoid misclassification of each sample in the dataset. By choosing the optimal value of parameter c , the proportion of errors in the determination of the solution will be small. The gamma parameter indicates how much curvature is desired within the decision limit [30].

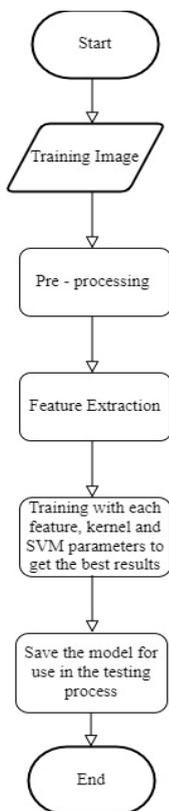


FIGURE 4. SVM Training Flowchart

Polynomial kernel is a kernel that is used when data cannot be separated linearly [28]. To perform analysis with this kernel, the cost (c) and degree (d) parameters are optimized. The parameter degree (d) serves to find the optimal value for each training sample. The greater the degree value, the resulting system will fluctuate because it will affect the

curvature of the resulting hyperplane line so that the selection of the degree value must be optimal [30]. The classification process in machine learning has two processes, namely the training process and the testing process. FIGURE 4 shows flowchart of training process, it begins with the mammogram image input process. The training image then goes through the pre-processing stage which aims to change the image size so as to save computational time. Furthermore, through the stage of extracting texture features from the training image. Texture feature extraction uses two methods, namely the GLCM method and the first order method. After obtaining the information data obtained from the feature extraction, the next step is the training stage using SVM. SVM training by determining the type of kernel to be used, which consists of Gaussian and Polynomial. The results of the training will be stored in a database which will later be used as a comparison in the testing process.

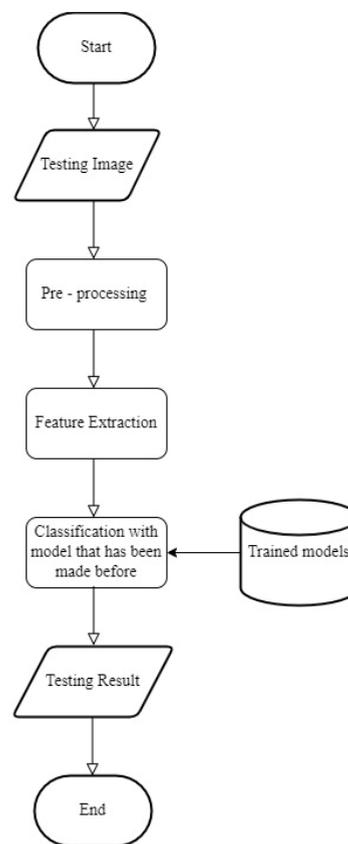


FIGURE 5. SVM Testing Flowchart

The testing process shows in FIGURE 5, it begins with the input of the test image which then goes through the pre-processing stage to shorten the computation time. The next step is feature extraction using the GLCM method and first order. After getting the feature value, the test image is identified using the SVM method to determine the suitability of the features or characteristics of the test data with the training data so that the results of the classification will be obtained.

III. RESULT

In this study using Matrix Laboratory (MATLAB) software by displaying a GUI design. There are several parameters to be tested by the system. The first test is the effect of changes in the pixel distance parameter on the Gray Level Co-Occurrence Matrix (GLCM). The images used consist of 400 training images, including 200 benign training images and 200 malignant training images, and 100 test images consisting of 50 benign test images and 50 malignant test images. In this test, feature extraction parameters were tested on GLCM with angles of 0°, 45°, 90°, and 135° with distance variations of 1, 2, and 3. The GLCM features used are contrast, correlation, energy, and homogeneity. Each feature will be tested so that the best features are obtained which will be selected for further testing. Then testing is done by combining all the features in GLCM. The test of the pixel distance aims to find out what the best neighbor distance value is needed by the system so as to obtain optimal results.

TABLE 2
SVM Accuracy to Distance with Every GLCM Feature

Pixel Distance	Accuracy			
	Contrast	Correlation	Energy	Homogeneity
1	54%	63%	63%	63%
2	59%	60%	64%	72%
3	67%	60%	63%	74%

Based on TABLE 2, there are the results of accuracy with the SVM method from each feature contrast, correlation, energy, and homogeneity. Testing with SVM uses Gaussian kernel type with parameter C 100 and gamma 1. The lowest accuracy is obtained with the contrast feature type with a distance of 1 pixel, which is 54%. The contrast value in each benign and malignant class has a value that is not much different so it is not good at distinguishing between benign and malignant classes. The best accuracy is obtained with the homogeneity feature with a distance of 3 pixels, which is 74% so that it is able to distinguish between benign and malignant classes better than other features.

TABLE 3
SVM Accuracy to Distance with GLCM Feature Combination

Pixel Distance	Accuracy
1	87%
2	88%
3	89%

Based on TABLE 3, shows the accuracy results obtained from the combination of all features of contrast, correlation, energy, and homogeneity with the SVM method. The best accuracy is obtained with the neighboring pixel distance of 3 pixels, which is 89%. Testing each pixel distance obtains an accuracy value past the target of 80%. In second test, the statistical feature type is tested on the first order by using the

mean, entropy, variance and skewness features. Each feature will be tested so that the best features are obtained which will be selected for further testing. Then the test is carried out by combining all the features on the first order.

TABLE 4
SVM Accuracy with Every First Order Feature

No.	First Order Features	Accuracy
1.	Mean	57%
2.	Entropy	62%
3.	Variance	64%
4.	Skewness	46%

TABLE 4 shows the results of the accuracy with the SVM method from each of the mean, entropy, variance and skewness features. Testing with SVM uses Gaussian kernel type with parameter c = 100 and value =1. The lowest accuracy is obtained with the variance feature type, which is 54%. The variance value shows the level of heterogeneity of the image histogram value. The best accuracy is obtained with the type of entropy feature, which is 62%. The entropy value indicates the level of randomness of the pixels in the image.

Then the test is carried out by combining all the features used in the first order, namely the mean, entropy, variance and skewness features with the SVM method. Accuracy is obtained that is equal to 51%. The resulting system obtained poor accuracy and did not exceed the target accuracy of more than 80%. In the third test, it is done by testing the parameters in the SVM kernel. The kernels used are Gaussian and Polynomial. There are two parameters to be tested with the Gaussian kernel type, namely the value of C and the value of gamma. Testing is carried out with the best features obtained from previous tests. By choosing the optimal value of parameter c, the proportion of errors in the determination of the solution will be small. The gamma parameter indicates how much curvature is desired within the decision limit [30].

TABLE 5
Accuracy of Parameter Testing on Gaussian Kernel With Homogeneity Features

No.	Parameter C	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
1.	$C = 1$	67%	61%	61%
2.	$C = 10$	73%	67%	65%
3.	$C = 100$	74%	75%	73%

TABLE 5 is the result of the accuracy of the test on the value of c and the value of gamma. The tested c values are 1, 10 and 100. While the gamma values tested are 1, 2, and 3. The test is carried out using the best feature in GLCM obtained from the first test, namely the homogeneity feature with a pixel distance of 3 pixels. The best accuracy is obtained with the value of C 100 with gamma 2 which is 75%, while the lowest accuracy is obtained using the value of C 1 with values of gamma 2 and 3. So in the test with the

homogeneity feature, the value of C 100 with gamma 2 is the optimal parameter.

TABLE 6

Accuracy of Parameter Testing on Gaussian Kernel With Combination of GLCM Features

No.	Parameter C	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
1.	$C = 1$	73%	68%	66%
2.	$C=10$	87%	82%	76%
3.	$C=100$	89%	89%	89%

TABLE 6 shows the accuracy result of testing the c value and gamma value by combining all GLCM features with a pixel distance of 3 pixels. namely contrast, correlation, energy, and homogeneity. The best accuracy obtained is 89% with a value of C 100 and gamma 1, 2, 3. The test results with a value of C 100 using either gamma 1, 2, or 3 each get the same accuracy value. While the lowest accuracy is obtained using the value of C 1 with a value of gamma 3. So in the test by combining all GLCM features with parameters C 100 with gamma 1, 2, and 3 is the optimal parameter.

TABLE 7

Accuracy of Parameter Testing on Gaussian Kernel With Variance Features

No.	Parameter C	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
1.	$C = 1$	50%	58%	56%
2.	$C=10$	54%	58%	58%
3.	$C=100$	54%	58%	58%

Tests on parameter C and gamma values were carried out with first-order feature extraction using the best feature in the previous test, namely the variance feature. Based on TABLE 7, tests with a gamma of 2 obtained the same accuracy of 58% either by using C 1, 10, or 100. Then the same accuracy was obtained with the parameter values of C 10 and 100 with a value of gamma 3. The lowest accuracy was obtained using the value of C 1 with a gamma of 1 that is 50%.

TABLE 8

Accuracy of Parameter Testing on Gaussian Kernel With Combination of First Order Features

No.	Parameter C	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
1.	$C = 1$	49%	49%	50%
2.	$C=10$	51%	49%	51%
3.	$C=100$	51%	49%	51%

TABLE 8 is the result of accuracy from testing the combination of all features on the first order, namely the mean, entropy, variance and skewness features. The highest accuracy was obtained with parameters C 10 and 100 with gamma values 1 and 3, namely 51%. While the lowest accuracy is 49% with gamma 2 using either a C value of 1, 10 or 100. Then testing using a Polynomial kernel type by testing the parameters C and degree. Degree affects the curvature of the resulting hyperplane line so that the selection of the degree value must be optimal.

TABLE 9

Accuracy of Parameter Testing on Polynomial Kernel With Homogeneity Features

No.	Parameter C	$d = 2$	$d = 3$	$d = 4$
1.	$C = 1$	67%	73%	72%
2.	$C=10$	74%	73%	74%
3.	$C=100$	74%	74%	81%

TABLE 9 shows the accuracy of the test on the value of c and degree. The tested c values are 1, 10 and 100. While the tested degrees are 2, 3 and 4. The test is carried out using the best features in GLCM obtained from the first test, namely the homogeneity feature with a pixel distance of 3 pixels. The best accuracy is obtained with a value of C 100 with degree 4 which is 81%, while the lowest accuracy is obtained using a value of C 1 with degree 2. So in the test with homogeneity features, the value of C 100 with degree 4 is the optimal parameter.

TABLE 10

Accuracy of Parameter Testing on Polynomial Kernel With Combination of GLCM Features

No.	Parameter C	$d = 2$	$d = 3$	$d = 4$
1.	$C = 1$	85%	89%	89%
2.	$C=10$	88%	89%	87%
3.	$C=100$	88%	86%	87%

TABLE 10 is the result of accuracy from testing the values of c and degree by combining all GLCM features, namely contrast, correlation, energy, and homogeneity. The best accuracy is obtained by 89% with degrees 3 and 4. With degree 3 the best accuracy is obtained using the values of C 1 and 10, while with degree 4 the highest accuracy is obtained with a value of C 1. While the lowest accuracy is obtained using the value of C 1 with degree 2. So in the test by combining all GLCM features using degrees 3 and 4 is the optimal value.

TABLE 11

Accuracy of Parameter Testing on Polynomial Kernel With Variance Features

No.	Parameter C	$d = 2$	$d = 3$	$d = 4$
1.	$C = 1$	54%	50%	50%
2.	$C=10$	54%	50%	50%
3.	$C=100$	54%	50%	50%

Tested on parameter C and degree value with first-order feature extraction using the best feature in the previous test, namely the variance feature. Based on TABLE 11 the highest accuracy produced is 54% with degree 2 using either C 1, 10 or 100 values. Tests on parameter C and degree values were carried out by combining the four features of the first order, namely the mean, entropy, variance and skewness features. TABLE 12 shows the highest accuracy is obtained with degree 2. The test with degree 2 obtains the same accuracy, both using C values 1, 10, and 100. Using degrees 3 and 4

produces the same accuracy, namely 50% with C values 1, 10, nor 100.

TABLE 12

Accuracy of Parameter Testing on Polynomial Kernel With Combination of First Order Features

No.	Parameter C	d = 2	d = 3	d = 4
1.	C = 1	64%	50%	50%
2.	C = 10	64%	50%	50%
3.	C = 100	64%	50%	50%

In the fourth test, combination of first order feature extraction and GLCM is performed by adjusting the distance of neighboring pixels in GLCM, which is 3 pixels because this distance is the optimal distance according to the results obtained in the first test. The total features used are 8 features taken from each of the GLCM and first order methods. Testing with the SVM method using the Gaussian kernel type with parameter values of C 100 and gamma 1 obtained a poor accuracy of 51%.

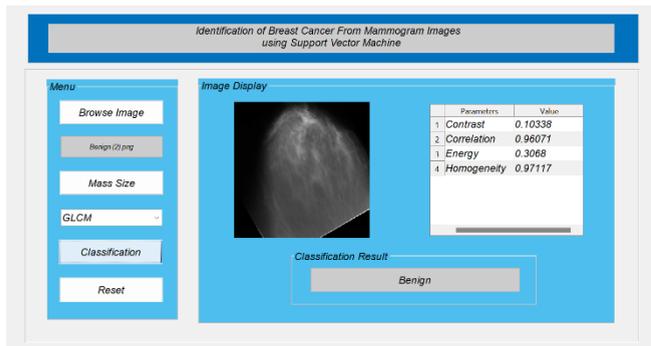


FIGURE 6 GUI Display

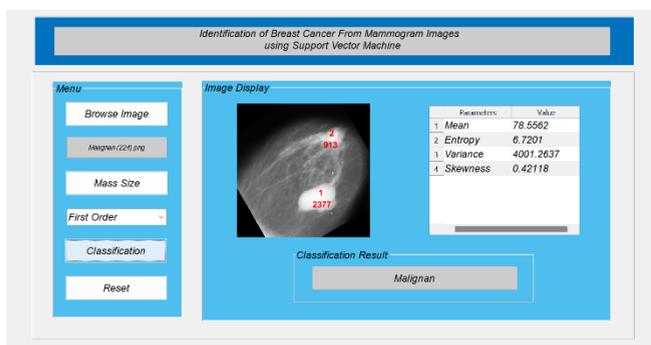


FIGURE 7. GUI Display

In this study using MATLAB software by displaying a GUI design. The GUI display has several features, including an browse input feature to select the test image to be classified. The image will be displayed on the Image Display panel. The feature extraction in the GUI uses a pop-up menu so that can be chosen to use the GLCM or first order method. The feature extraction results will be displayed in the table. After getting the feature extraction value, then the classification is carried out using a button named 'Classification'. The results

of the classification will be displayed on the 'Classification Results' menu. Reset menu to clear the acquired information and return to the initial stage. The following is the display form of the GUI on this system, shown in FIGURE 6 and FIGURE 7.

IV. DISCUSSION

After several experiments, obtained different accuracy values. The first test is to test the variation of pixel spacing in GLCM. The test is carried out with each GLCM feature obtaining the highest accuracy with a distance of 3 pixels with the homogeneity feature and the lowest accuracy is obtained with the contrast feature. Homogeneity measures the degree of homogeneity or similar variations in gray intensity image, which is the opposite of the contrast. So that means if the homogeneity increases, the contrast decreases [15][31]. At a distance of 3 pixels produces good feature extraction so that it can distinguish between the two classes. This is because the pixel pairs with a distance of 3 pixels have a lot of uniformity of appearance. Further testing is done by combining the four GLCM features, the best results are obtained using SVM, which is at a distance of 3 pixels. The accuracy results obtained by combining all the features in GLCM obtain much better accuracy than using only 1 statistical feature. This is because all these features are interconnected with each other. Compared to previous research, only using the default pixel spacing, which is 1 pixel, while in this study testing different pixel spacings so that the best results are obtained, namely at a distance of 3 pixels.

Then the second test using the first order method. The accuracy results obtained low values, both testing using only one statistical feature or combining all four features. By combining four features, the highest accuracy is only able to reach 51%. In the third test, the Gaussian and Polynomial kernel types were tested by changing each parameter. The selection of the optimal C parameter is very influential on the system so that it can reduce misclassification between classes. Then the gamma parameter in the Gaussian kernel affects the level of accuracy so it is necessary to choose the optimal value. In this study, the optimal C values were obtained, namely C 100 and gamma 1, 2, and 3 with GLCM feature extraction. Then in the polynomial kernel there is a degree parameter that determines the curvature of the hyperplane line. In this study, the optimal degree values are 2 and 3 and the value C 1 and 10 with GLCM feature extraction. Compared to previous studies, it's not setting the best hyperparameters and only choosing random parameters used in SVM. Then testing is also carried out with the variance feature and the combination of the four features on the first order. This test obtained a low accuracy value with an accuracy below 80%. So a system using GLCM feature extraction is a better method than first order because GLCM produces much higher accuracy.

Seen from the results obtained from the experiment, the highest accuracy only reached 89% and was not able to reach 100%, so there is still a possibility that the system can predict incorrectly and mistakenly distinguish classes. Adding a quality improvement process to the preprocessing process can be an alternative to improve accuracy by reducing noise in the dataset.

IV. CONCLUSION

The breast cancer identification system goes through several stages, namely the mammogram image input process, then pre-processing is carried out to resize the image size. Then the texture feature extraction stage with the GLCM method and first order. After getting the feature results, then the classification is carried out using the SVM method.

The purpose of this study is to compare two types of texture feature extraction methods, namely the GLCM method and the first order using the SVM classification method in order to obtain a system with the best accuracy that can distinguish between the two benign and malignant classes. The results showed accuracy using the SVM method with polynomial and Gaussian kernels by combining the four GLCM features obtains the best accuracy value of 89%. The best results with the polynomial kernel use C 1 and 10 with degree 3 and 4, while using the Gaussian kernel use C 100 with gamma 1, 2, and 3. So the SVM classification method with GLCM feature extraction is the best method for identifying breast cancer using mammogram images. For future studies, at the pre-processing stage an image quality enhancement process can be added to reduce noise that can make predictions wrong so as to obtain the highest accuracy up to 100%.

REFERENCES

- [1] R. I. Kemenkes, "Kasus Kanker Payudara Paling Banyak Terjadi di Indonesia," *Databoks*, p. 2019, 2019.
- [2] The Global Cancer Observatory, "Cancer Incident in Indonesia," *Int. Agency Res. Cancer*, vol. 858, pp. 1–2, 2020, [Online]. Available: <https://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf>
- [3] S. M., A. A., H. E., and M. T., "Breast Cancer Detection with Mammogram Segmentation: A Qualitative Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 117–120, 2017, doi: 10.14569/ijacsa.2017.081016.
- [4] J. Poorolajal, M. E. Akbari, F. Ziaee, M. Karami, and M. Ghoncheh, "Breast cancer screening (BCS) chart: A basic and preliminary model for making screening mammography more productive and efficient," *J. Public Heal. (United Kingdom)*, vol. 40, no. 2, pp. e118–e125, 2018, doi: 10.1093/pubmed/idx052.
- [5] M. Löberg, M. L. Lousdal, M. Bretthauer, and M. Kalager, "Benefits and harms of mammography screening," *Breast Cancer Res.*, vol. 17, no. 1, pp. 1–12, 2015, doi: 10.1186/s13058-015-0525-z.
- [6] S. Saini and R. Vijay, "Mammogram analysis using feed-forward back propagation and cascade-forward back propagation artificial neural network," *Proc. - 2015 5th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2015*, pp. 1177–1180, 2015, doi: 10.1109/CSNT.2015.78.
- [7] S. Charan, M. J. Khan, and K. Khurshid, "Breast cancer detection in mammograms using convolutional neural network," *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCOMET 2018 - Proc.*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ICOMET.2018.8346384.
- [8] A. Karahaliou *et al.*, "Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis," *Br. J. Radiol.*, vol. 80, no. 956, pp. 648–656, 2007, doi: 10.1259/bjr/30415751.
- [9] N. M. Basheer and M. H. Mohammed, "Classification of Breast Masses in Digital Mammograms Using Support Vector Machines," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 10, p. 2277, 2013, [Online]. Available: www.ijarcsse.com
- [10] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Meas. J. Int. Meas. Confed.*, vol. 146, pp. 800–805, 2019, doi: 10.1016/j.measurement.2019.05.083.
- [11] T. Lin and M.-L. Huang, "Dataset of Breast mammography images with Masses," 2020.
- [12] S. Ahmed Medjahed, "A Comparative Study of Feature Extraction Methods in Images Classification," *Int. J. Image, Graph. Signal Process.*, vol. 7, no. 3, pp. 16–23, 2015, doi: 10.5815/ijigsp.2015.03.03.
- [13] D. P. Tian, "A review on image feature extraction and representation techniques," *Int. J. Multimed. Ubiquitous Eng.*, vol. 8, no. 4, pp. 385–395, 2013.
- [14] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019, doi: 10.1109/ACCESS.2018.2890743.
- [15] M. Hall-Beyer, "GLCM Texture: A Tutorial vol 3.0," *Creat. Commons Attrib. 4.0 Int. Licens.*, no. February, 2017, [Online]. Available: <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm>
- [16] P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar, "Image Texture Feature Extraction Using GLCM Approach," *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, pp. 290–294, 2013.
- [17] N. Zulpe and V. Pawar, "GLCM textural features for Brain Tumor Classification," *Int. J. Comput. Sci.*, vol. 9, no. 3, pp. 354–359, 2012, [Online]. Available: <http://www.doaj.org/doi?func=abstract&id=1158398>
- [18] A. Ehsanirad and S. K. Y. H., "Leaf recognition for plant classification using GLCM and PCA methods," *Orient. J. Comput. Sci. Technol.*, vol. 3, no. 1, pp. 31–36, 2010.
- [19] M. Wati, Haviluddin, N. Puspitasari, E. Budiman, and R. Rahim, "First-order Feature Extraction Methods for Image Texture and Melanoma Skin Cancer Detection," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019, doi: 10.1088/1742-6596/1230/1/012013.
- [20] M. Rajalakshmi and P. Subashini, "Texture Based Image Segmentation of Chili Pepper X-Ray Images Using Gabor Filter," *Int. J. Adv. Stud. Comput. Sci. Eng. IJASCSE*, vol. 3, no. 3, pp. 44–51, 2014, [Online]. Available: www.ijascse.org
- [21] N. Aggarwal and R. K. Agrawal, "First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images," *J. Signal Inf. Process.*, vol. 03, no. 02, pp. 146–153, 2012, doi: 10.4236/jsip.2012.32019.
- [22] L. H. Thai, T. S. Hai, and N. T. Thuy, "Image Classification using Support Vector Machine and Artificial Neural Network," *Int. J. Inf. Technol. Comput. Sci.*, vol. 4, no. 5, pp. 32–38, 2012, doi: 10.5815/ijitcs.2012.05.05.
- [23] B. C. Kuo, H. H. Ho, C. H. Li, C. C. Hung, and J. S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 1, pp. 317–326, 2014, doi: 10.1109/JSTARS.2013.2262926.
- [24] S. Agrawal, N. K. Verma, P. Tamrakar, and P. Sircar, "Content based color image classification using SVM," *Proc. - 2011 8th Int. Conf. Inf. Technol. New Gener. ITNG 2011*, pp. 1090–1094, 2011, doi: 10.1109/ITNG.2011.202.
- [25] X. Zhou, X. Zhang, and B. Wang, "Online support vector machine: A survey," *Adv. Intell. Syst. Comput.*, vol. 382, no. 8, pp. 269–278, 2016, doi: 10.1007/978-3-662-47926-1_26.
- [26] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," *Sensors (Switzerland)*, vol.

- 12, no. 9, pp. 12489–12505, 2012, doi: 10.3390/s120912489.
- [27] A. E. Hassanian and T. H. Kim, “Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks,” *J. Appl. Log.*, vol. 10, no. 4, pp. 277–284, 2012, doi: 10.1016/j.jal.2012.07.003.
- [28] D. R. Tobergte and S. Curtis, *Learning with Kernels*, vol. 53, no. 9. 2013.
- [29] I. Sitorus, “Support Vector Machine (SVM) and Kernels Trick,” 2020. <https://medium.com/analytics-vidhya/introduction-to-svm-and-kernel-trick-part-1-theory-d990e2872ace>
- [30] I. Sitorus, “Support Vector Machine (SVM) and Kernels Trick,” *Analytics Vidhya*, 2020. <https://medium.com/analytics-vidhya/introduction-to-svm-and-kernel-trick-part-1-theory-d990e2872ace>
- [31] A. A. Gade and A. J. Vyavahare, “Feature Extraction using GLCM for Dietary Assessment Application,” *Int. J. Multimed. Image Process.*, vol. 8, no. 2, pp. 409–413, 2018, doi: 10.20533/ijmip.2042.4647.2018.0050.