Similarity measurement on digital mammogram classification

Erna Alimudin[1], Hanung Adi Nugroho[2], Teguh Bharata Adji[2]

[1]Study Program of Electronics Engineering, Department of Electronics Engineering, Polytechnic State of Cilacap, Center Java, Indonesia

[2]Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Article history:

Received Jun 13, 2020
Revised Jun 16, 2022
Accepted Jun 25, 2022

Keywords:

EBP
GLCM
Histogram
Mammogram
Similarity measurement

ABSTRACT

Breast cancer is one of the dominant causes of death in the world. Mammography is the standard for early detection of breast cancer. In examining mammograms, the overall parenchyma pattern of the left and right breast was placed side by side for symmetry assessed of left and right breast tissue by radiologist. Thus, in building computer-aided diagnosis (CAD) system for screening mammography, it is necessary to adapt the working procedure of the radiologist. In this study, 30 training images and 30 testing images from Kotabaru Oncology Clinic in Yogyakarta were used. The first step was to enhance the image quality with median filter and contrast limited adaptive histogram equalization (CLAHE). Then, feature extraction was processed by histogram-based and by gray level co-occurrence matrix (GLCM) based. Furthermore, the similarity measurement process was used to measure the difference value between selected features, i.e. angular second moment (ASM), inverse difference moment (IDM), contrast, entropy based GLCM, and energy, on the left and right mammograms. This process was intended to assess the symmetry of left and right mammograms as radiologists do in mammography screening. The obtained results of the classification between normal and abnormal images with backpropagation algorithm were accuracy of 0.933, sensitivity of 0.833, and specificity of 1.000.

Corresponding Author:

Erna Alimudin
Study Program of Electronics Engineering, Department of Electronics Engineering
Polytechnic State of Cilacap, Dokter Soetomo Road, No 1, Cilacap District 53212
Center Java, Indonesia
Email: ernaalimudin@pnc.ac.id

1. INTRODUCTION

The International Agency for Research on Cancer (IARC) is an exclusive cancer research agency of World Health Organization (WHO). The Global Cancer Observatory (GLOBOCAN) is a project of IARC. GLOBOCAN is an interactive web-based platform presenting statistical data of to give information about cancer control and research. GLOBOCAN presents statistical data based on estimation from cancer sites and sex using the best available data in each country and several estimation methods. The IARC published the latest estimates on the global burden of cancer on September 2018. Lung, female breast, and colorectal cancer are the three types of cancer that have the highest incidence, and are in the top five in mortality rates (first, fifth, and second, respectively). One third of the cancer incidence and the burden of death in the world is the incidence of these three types of cancer [1]. Overall, in recent years several Asian countries have inspected a significant increase of breast cancer in the incidence with the incidence rate increasing by 3% to 4% per year in China, Singapore and Thailand [2]. Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among women and the leading cause of cancer death [3].

Number of breast cancer new cases in South Eastern Asia (Brunei, Myanmar, Cambodia, Indonesia, Laos, Malaysia, the Philippines, Singapore, Thailand, and Vietnam) in 2020 by all ages of women is 158,939 (27.7%) of a total 573,847 female cancer cases. This is the highest cases compared to other common cancer i.e. lung, color-rectum, ovary, and cervix uteri [4]. Number of breast cancer new cases in Indonesia in 2020 by all ages of women is 65,858 (30.8%) of a total 213,546 female cancer cases. This is the highest cases compared to other common cancer i.e. cervix uteri, ovary, color-rectum, and thyroid [5].

The international incidence of female breast cancer has been estimated to hit approximately 3.2 million new cases per year by 2020 [6]. In addition to population growth, breast cancer incidence rates are projected to rise further in many less developed countries due to longer life expectancy coupled with the adoption of a more "westernized" lifestyle and increased population-based screening. As a result, the global burden of breast cancer in the Asian region is expected to be strongly affected by improvements in incidence [7]. Therefore, to reduce such cases in the future, breast cancer must be detected at its early stage through a proper screening. Screening that refers to finding symptoms. The goal of screening is to find cancer [7]. The earlier cancer is detected, the greater probability of successful treatment of the disease.

Mammography and ultrasonography are standard breast diagnoses [8]. Mammography is a special radiological examination using low-dose X-rays to detect abnormalities in the breast. Mammography is a way for early detection and the effective technique for breast cancer screening [9]. X-ray images from mammography screening called as mammogram. X-rays will reduce the thickness of breast tissue and hold the breast position by compressing it, thus giving the radiologist the ability to read the disorder more clearly [10].

After screening mammography, mammogram will be interpreted by the radiologist [11]. The interpreting radiologist must be an expert and experienced [12]. Radiologists cultivate hundreds of mammograms everyday so it is hard to maintain the consistency and precision of diagnosis. A radiologist might just miss some of the disorder [9], [10]. Therefore, to read the mammogram, two radiologist should be presented [12], [13]. Differences of opinion can occur between the two radiologists, for example a radiologist detects abnormalities while the other does not, then a mammogram will be re-assessed by a third reader [14], [15]. A misinterpretation and incorrect result of mammogram screening, will cause an increase in mortality rate per incidence. In regards to these matters, an automation system is needed to be used as the third reader tools that can help the radiologist to interpret the mammogram. One such system is computer aided diagnosis (CAD). CAD systems are a substantial way to detect breast cancer and reduce disease morbidity [13], [14].

CAD on mammogram images has been studied in previous studies using various similarity measurement methods but they had the same purpose. The purpose is classification of mammograms [16], [17]. Singh *et al*. [18] has also done other research on mammogram image classification. In this work, normal and abnormal images were classified using random forest classifier and gray level co-occurrence matrix (GLCM) based informative texture features, where informative features were selected using Ada-Boost feature selection method. This work was continued by using a content-based image retrieval (CBIR) to classify normal and abnormal mammogram images. Singh *et al*. [19] performed CBIR by classifying mammogram queries and taking similar mammograms already described by the diagnostic description and treatment results. The final step, classification was determined by discovering similar images which selected using the Euclidean distance similarity measure. Mammogram images were compared from the benchmark. Thus, in this study, a comparison was made between the mammogram image of a patient and the mammogram mammographic images analysis society (MIAS) database. There were no comparisons between the left and right mammogram image of a patient as the radiologist do to observe mammography screening result. This study obtained the effectiveness of the proposed work regarding an average precision of 72% for classifying normal mammogram images and 61.30% for classifying abnormal mammogram images. Setiawan *et al*. [20] proposed a mammogram classification using an artificial neural network (ANN). The MIAS database was used as training data for the mammogram classification model taken from. In this research, in addition to LAW's texture, GLCM was also tried to be used as feature extraction. The accuracy result was 72.20% for normal-abnormal classification. Liu [21] also did similarity measurement step with the Euclidean distance measurement. The data consisted of 30 benign and 30 malignant cancer samples were trimmed at the ROI then extracted the features with GLCM. The study produced a result on about 58% accuracy.

Several previous studies showed different step of CAD-based research. The value of accuracy obtained from previous studies is relatively low. Therefore, it can be concluded that the CAD method is ineffective. After being observed, it was found that the stage of similarity measurement carried out was not in accordance with the procedures carried out by the radiologist [22]. In previous studies, images were classified by similarity measurements using CBIR [23]. The CBIR principle is to determine the testing of images with the same features [24]. However, this is quite contrary to the actual procedure, because each patient's breast characteristics are not same, depending on the patient's age and condition. For example, at a certain age, pregnant or breastfeeding women, or women who have a dense breast texture, can be categorized as abnormal when viewed from one side only, whereas when viewed from two sides, the patient's breasts look symmetrical and can be declared normal [25]. In carrying out similarity measurements, the radiologist compares the right mammogram and left mammogram to observe abnormalities [16] it is based on guideline. This guide was created collaboratively by individuals with proficiency in breast imaging, medical physics, and imaging informatics, representing the American College of Radiology (ACR), the American Association of Physicists in Medicine (AAPM), and the Society for Imaging Informatics in Medicine (SIIM), especially for technical guidance. One of the imaging tasks on mammography to visualize the following features of breast cancer is the asymmetry observation between left and right breast images [26]. The cranio-caudal (CC) view should expose as much of the breast as possible. A correctly performed cc view will show nearly all the breast except the most lateral and axillary part. One of the criteria for assessing the cc view is symmetrical images [27]. The illustration is shown in Figure 1, this is done according to the results that someone will have a mammogram with the same characteristics between left and right. Thus, the different stages of similarity measurement between previous studies and radiological procedures are problems that need to be solved.

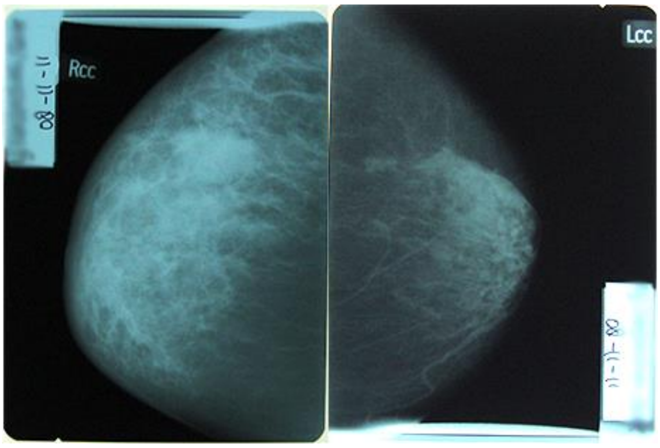


Figure 1. The right mammogram and left mammogram compared to observe abnormalities

The different stages of similarity measurement between previous studies and radiological procedures influence the decision of classification result. The decision of classification result in previous research was not supported by the observation result of left and right mammogram as radiologist did in mammography screening. Therefore, in this study, the low accuracy value that was obtained from previous studies could be solved. This study follows the similarity measurements carried out by radiologists in observing the results of mammography screening. It is conducted on the similarity measurement step by measuring the similarity between left and right side of digital mammogram features of patients.

Mammogram should be inspected in optimal lighting situations. Films should be checked whether the label identity is valid and the radiographic quality should be relevant. The whole pattern of breast parenchyma is evaluated. Standard of medio-lateral oblique (MLO) and CC images projection are studied correctly on the left and right films ‘back-to-back’ thus allowing the symmetry of left and right breast tissue to be checked. A systematic search for signs of abnormal mammography is made and any abnormal signs should be analyzed to determine the need for other screening examinations [28].

2. RESEARCH METHOD

The data provided by the Kotabaru Oncology Clinic in Yogyakarta were mammogram images which were the result of patient mammography data. Each patient’s mammogram images consisted of two images, left and right breast, which were taken from CC point of view, named right craniocaudal (RCC) and left craniocaudal (LCC). The mammogram image was an X-ray film with a large size of 4,000×3,000 pixels which was digitized with a digitizer. It will require a lot of computing time during pre-processing. Also, the digitized image contains other information components about the patient that were scanned during the digitization process. Thus, the cropping process must be carried out to remove the patient identity in order to maintain the confidentiality of patient data and eliminate parts of the image that do not contain the information needed. Moreover, cropping process can be ended by compressed the image for the computational process efficiency. The image was cropped into size of 1,400 by 1,850 pixels and compressed into tag image file format (TIFF). TIFF was chosen because TIFF uses lossless compression to maintain integrity and clarity of the image [29]. The whole images were inputted and converted into a grayscale format.

The data were divided into two purposes in this research, i.e. training data and testing data which each contains of 10 normal images and 20 abnormal images. Total data used in this research were 60 mammogram images of patients. The given data by the radiologists were divided into three folders based on the characterization of breast mass as the examinations result by radiologists, namely normal, benign, and malignant. Each folder contains 20 mammogram images. The most common abnormality that causes breast cancer was its masses. Breast masses severity can be categorized as benign and malignant [30]. Thus, benign and malignant images were classified as abnormal images in this research. There were 40 abnormal images which consist of 20 benign mammogram images and 20 malignant mammogram images.

This research was divided in two main stages, i.e. learning stage and testing stage. In the learning stage, the database of training data was created. Training data had been labeled as normal or abnormal as when saved in database. The flowchart of learning stage can be seen in Figure 2 and the flowchart of testing stage can be seen in Figure 3.

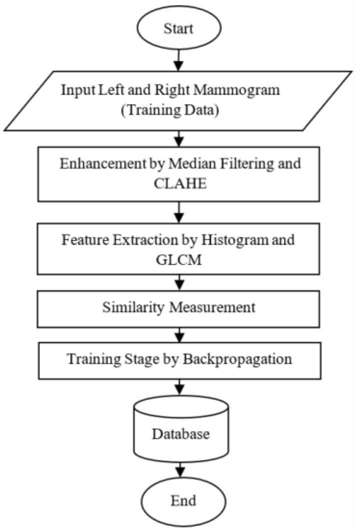


Figure 2. Learning stage flowchart

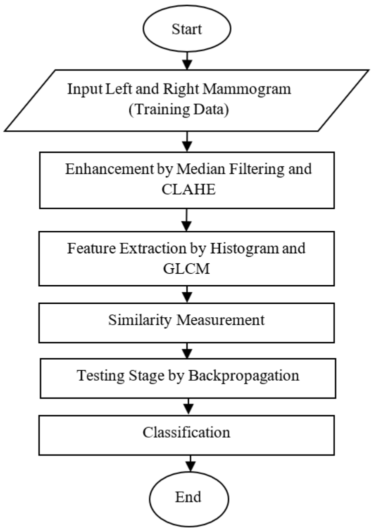


Figure 3. Testing stage flowchart

The learning stage and testing stage had six steps. The first steps until five steps were the same. The different step is only in the last step. The last step in learning stage was storing the processed value by backpropagation to the database. While, the last step in testing stage was to classify between normal or abnormal, which based on the database in the learning stage.

First step, left and right mammogram image of patient was inputted. The next step was image enhancement. Blur and noise are general features of undesirable elements in the medical image because it can reduce the visibility of certain objects. Median filter is able to reduce noise and to reduce blur. In addition, the physical contrast in the soft tissue chest is very low. Therefore, the contrast limited adaptive histogram equalization (CLAHE) method was used to overcome the problem of contrast. It is capable to improve the mammogram contrast image. Median filter also used to reduce noise and reduce blur thus the mammogram image is clearer. Both methods are able to increase image quality and meet the determinants of the radiography image quality [31].

The third step was feature extraction. It was a calculation process that produces a number of feature values of the mammogram image. The features were extracted by histogram-based features and GLCM to obtain the characteristics of the image. The number of features used does not always provide increased accuracy. Feature selection is a way to obtain relevant features which can actually increase accuracy. This is because the relevant features represent the image class. From the previous research, it is known that ASM, contrast, IDM, entropy-based GLCM, and energy, were relevant features which could obtain high accuracy [32], [33]. Thus, these five features as shown in Table 1 were used in our proposed approach.

The next step was the main idea of this research, which was similarity measurement. Similarity measurement for any two images is commonly obtained by measuring the distance between their extraction feature. This step adapted the procedure of radiologist when examining patients by comparing left and right mammogram images [24]. The distance represents the difference value of the left and right mammogram image feature extraction results of each patient. This step would measure the similarity between each feature of left and right mammogram. The calculations were carried using (1) [34].

$$symmetry = \frac{\Sigma i\ |left\ i - right\ i|}{\Sigma i\ |left\ i + right\ i|} \quad (1)$$

Normal patient will have a similarity measurement value which is very small. This value is obtained from left and right normal mammogram image feature values which are almost the same or symmetry. On the contrary, abnormal patients will have larger similarity measurement value. This value is obtained from one mammogram image, left or right side, which abnormal, thus the value will be different than the normal one. It is because left and right breasts of the patient are asymmetrical.

Table 1. Features used in phase extraction feature

Feature	Formula	Explanation
ASM	$\sum_{x=1}^{L}\sum_{y=1}^{L}(GLCM(x,y)^2$	Image homogeneity measurement
Contrast	$\sum_{n=1}^{L} n^2 \left\{\sum_{\|x-y\|=n} GLCM(x,y)\right\}$	Measurement of grayscale level pixel existence in image
IDM	$\sum_{x=1}^{L}\sum_{y=1}^{L}\frac{(GLCM(x,y)^2}{1+(x-y)^2}$	Homogeneity measurement
Entropy based GLCM	$-\sum_{x=1}^{L}\sum_{y=1}^{L}(GLCM(x,y)\log(GLCM(x,y)$	Measurement of grayscale level irregular in image
Energy	$\sum_{i=0}^{L-1}[p(i)]^2$	Measurement of pixel intensity distribution by grayscale level

Note:
i = gray level in the image
$p(i)$ = probability of emergence of i
L = the highest gray level value
x, y = coordinates (x, y) indicate the location/manner of pixels in an image

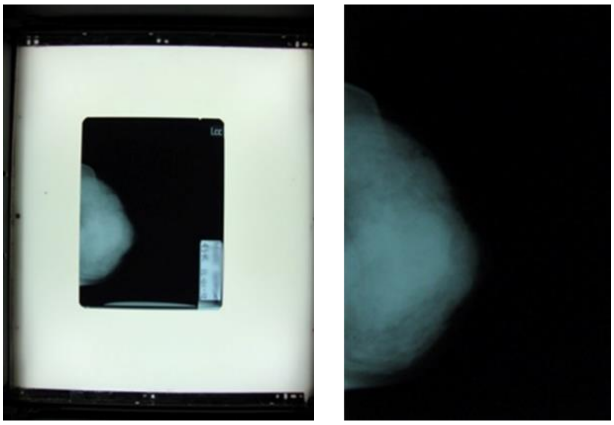
The next step was done by backpropagation algorithm. The learning stage was executed first to get the database needed as a reference for the testing stage. At the learning stage, these similarity measurement values were trained with backpropagation algorithm to obtain weight values that can be used in testing the data to be classified in testing stage. Backpropagation is a guided learning algorithm and is usually used by multi-layer perceptron (MLP) to change the weights associated with neurons in the hidden layer. The network was given a pattern consisting of an input of pattern from similarity measurement value of the mammogram images inputted and the desired pattern in the output from labeled mammogram images label which known as normal or abnormal. When a pattern was assigned to the network, the weights were changed to minimize differences in the output pattern and the desired pattern. This learning was repeated so that all the patterns released by the network can meet the desired pattern. This weight was then stored in the database and used for further testing processes. At the testing stage, the similarity measurement values from the testing data images will be used as backpropagation input. Testing data images were a number of mammogram images that have never been used as training data at the learning stage. Last step is to look for the value of neurons in the hidden layer and the output matches the pattern obtained from the learning results. Then, look for the error between the output neuron value with each target that has been normalized. Finally, at the testing stage, normal or abnormal class will be determined based on the smallest error value. After all stages of the research were completed, a system in the form of a graphic user interface (GUI) has been made to make it easier to use. The main purposed of this study was a classification system for mammography screening mammogram.

3. RESULTS AND ANALYSIS

The entire mammogram image used in this research was pre-processed. It was required to prepare the image for further process [35]. The entire mammogram images were trimmed to remove the patient label and was compressed lossless in TIFF format. Digitized mammogram image can be seen in Figure 4(a), and the results after trimmed and compressed TIFF can be seen in Figure 4(b).

After trimmed and compressed, the mammogram image was enhanced by median filter and CLAHE method. Mammogram image before enhancement process shown in Figure 5(a), after enhanced by median filter in Figure 5(b), and the result of mammogram image after enhanced by median filter and CLAHE in

Figure 5(c). The result showed that the contrast of the mammogram image was increased. Noise and blur are also reduced. Thus, making the mammogram image is clearer. Similarity measurement step in this research was conducted to adapt the work procedure of the radiologist in examining the results of mammography screening. It was done by calculating the difference value between the statistical features between left and right mammograms of normal patients and abnormal patients. The purpose is to measure the symmetry between left and right mammograms.



(a) (b)

Figure 4. Comparing digitized mammogram image: (a) before pre-processing and (b) after trimmed and compressed to TIFF format

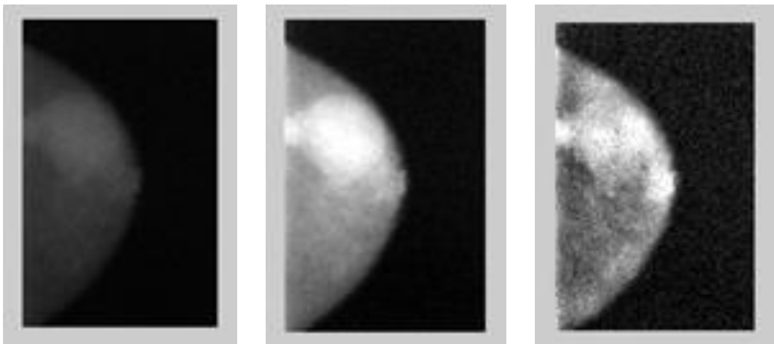
The given data by the radiologists was divided into three folders based on the characterization of breast mass as the examinations result by radiologists, namely normal, benign, and malignant. Each folder contains 20 mammogram images of patients. The most common abnormality that causes breast cancer was its masses. Breast masses severity can be categorized as benign and malignant [30]. Thus, benign and malignant images were classified as abnormal images in this research. There were 40 abnormal images which consist of 20 benign mammogram images and 20 malignant mammogram images. Total data used in this research were 60 mammogram images of patients. The data were divided into two purposes in this research, i.e. training data and testing data which each contains of 10 normal images and 20 abnormal images.

Similarity measurement value of each feature of average data used of normal, benign, and malignant (abnormal) can be seen in Table 2 and clearly showed by diagram in Figure 6 to Figure 10. Data used were 20 normal mammogram images and 40 normal mammogram images which consist of 20 benign mammogram images and 20 malignant mammogram images. Similarity measurement value of each feature of average data used of normal, benign, and malignant (abnormal) can be seen in Table 2 and clearly showed by diagram in Figure 6 to Figure 10. Data used were 20 normal mammogram images and 40 normal mammogram images which consist of 20 benign mammogram images and 20 malignant mammogram images.

Values in Table 2 and diagram in Figure 6 to Figure 10 shows how much asymmetry value between the left and right mammograms. It will be the parameter of abnormality existence on patient's mammogram. A normal patient has a similarity measurement value which is smaller than abnormal. It is because of both mammograms are normal that features value of left and right mammograms are about the same. On the contrary, an abnormal patient has similarity measurement value is larger than normal. It is because one of left and right mammogram is abnormal that the features values of left and right mammogram are not same. Thus, the value is big difference. It describes that the similarity measurement step is right to do. Similarity measurement step is done by measure symmetry between the left and right mammogram of patients. This step adapts the working procedure of radiologist in mammography screening when they observe the overall pattern left and right breast on mammogram X-ray results. The similarity measurement step was the key to the success of the classification step. Similarity measurement values were trained in learning stage by backpropagation. Backpropagation algorithm in this research used two hidden layers that previously had been trained with 30 training data which consisted of 10 normal mammogram images and 20 abnormal mammogram images. The pattern result of learning stage consisted of similarity measurement values, weight, layer, and the desired output as normal or abnormal. All of them were stored in database. Lastly, the classification step in testing stage used backpropagation method and worked according to the pattern result of learning stage.

Table 2. Similarity measurement for each feature in normal and abnormal mammogram

Average value of feature from 20 mammogram images	GLCM feature				Histogram feature
	ASM	Contrast	IDM	Entropy	Energy
Normal	0.0007	0.740649	0.0046	0.1322	0.0024
Abnormal (Benign)	0.0082	3.2507285	0.0407	0.3326	0.01435
Abnormal (Malignant)	0.0121	5.0014525	0.0618	0.5621	0.0192



(a) (b) (c)

Figure 5. Mammogram image: (a) before enhancement, (b) after enhanced by median filter, and (c) after enhanced by median filter and CLAHE

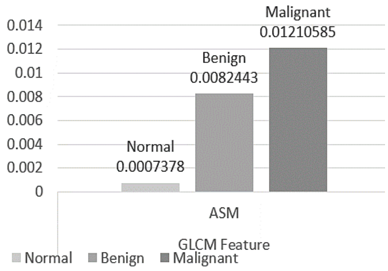


Figure 6. ASM value comparison diagram

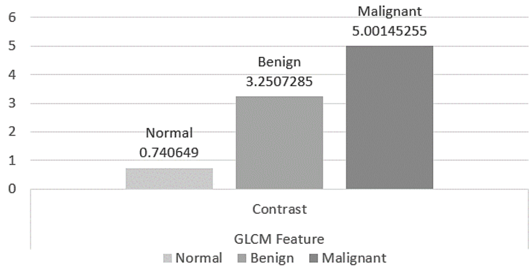


Figure 7. Contrast value comparison diagram

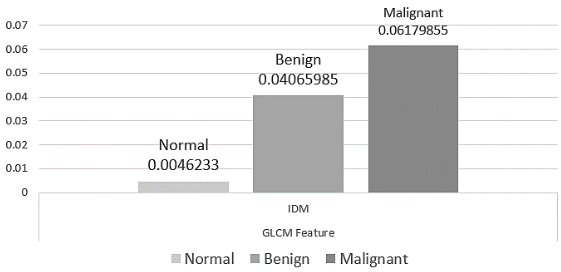


Figure 8. IDM value comparison diagram

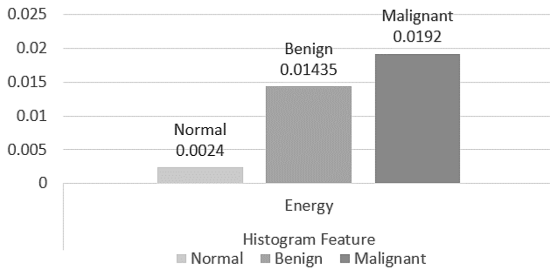


Figure 9. Energy value comparison diagram

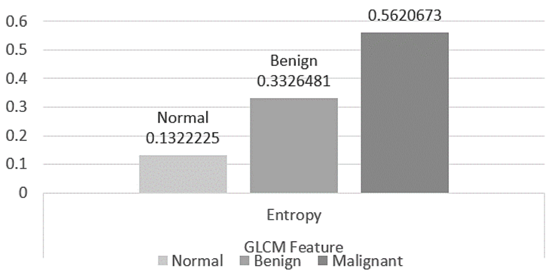


Figure 10. Entropy value comparison diagram

Accuracy, sensitivity, and specificity are scalar values in different metrics that can represent classification performance. Graphical assessment methods such as receiver operating characteristics (ROC) provide different interpretations of classification performance [36]. Accuracy (Acc) is a measure of classification performance obtained by defining the ratio between properly classified samples and the number of samples according to (2). The sensitivity, true positive rate (TPR), or recall of a classifier represents samples that were classified correctly positive to the total number of positive samples, and it is measured by using (3). While specificity, true negative rate (TNR), or precision is described as the ratio of correctly classified negative samples to the total number negative sample and it is measured by using (4). These classification metrics can be calculated based on the data extracted from the confusion matrix [37]. The confusion matrix of classification on 30 mammogram images. The testing data was 30 testing data which consisted of 10 normal mammogram images and 20 abnormal mammogram images that had not been trained before. The classification result showed by confusion matrix in Table 3. There were 18 abnormal mammogram images obtained and classified as abnormal mammogram images (true positive), 2 abnormal mammogram images classified as normal mammogram images (false negative), 0 normal mammogram image classified as abnormal mammogram image (false positive), and 10 normal mammogram images classified as normal mammogram images (true negative).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$TPR = \frac{TP}{TP+FN} = \frac{TP}{P} \tag{3}$$

$$TNR = \frac{TN}{FP+TN} = \frac{TN}{N} \tag{4}$$

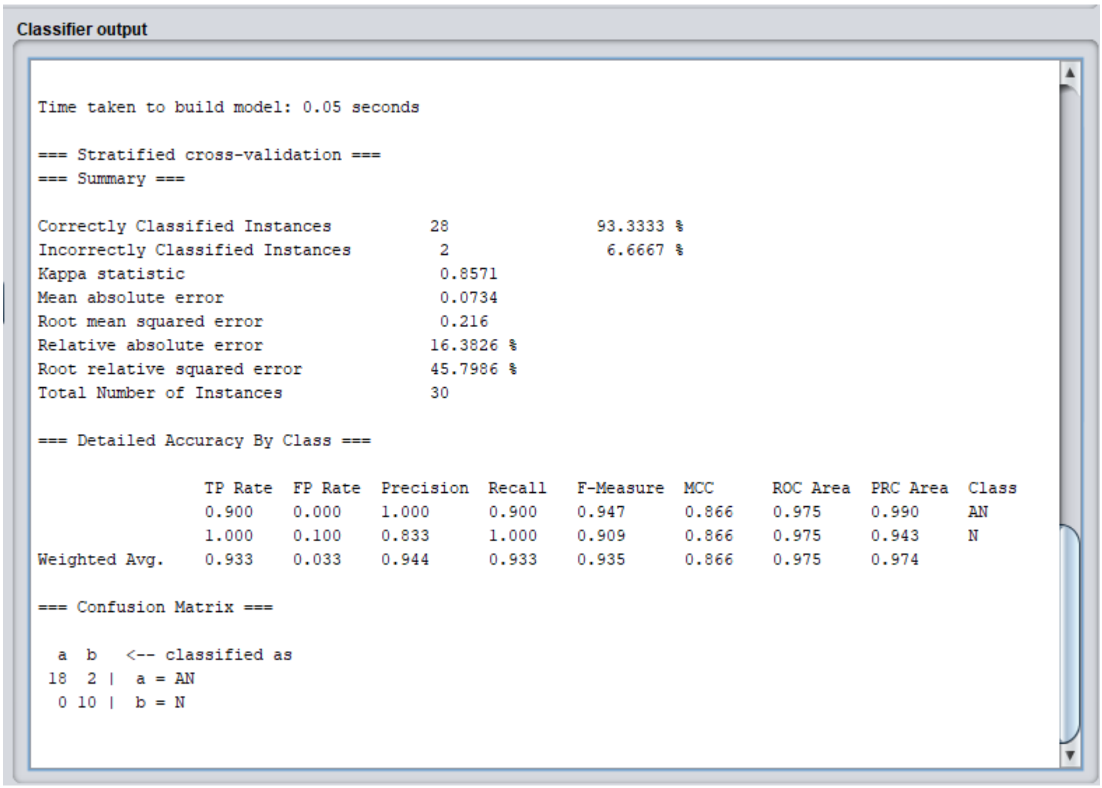
The detailed accuracy by class and the ROC was measured by using Weka 3.6 of 10 folds of cross validation, the result shows in Figure 11 and Figure 12. From the results, it was known the value of accuracy was 0.933, sensitivity was 0.900, specificity was 1.000. The result also obtained the matthews correlation coefficient (MCC) of 0.866, and the F-measure of 0.947. MCC represents the correlation between the observed and predicted classifications. While, F-measure is a measure of test accuracy. The result indicates accuracy, sensitivity and high specificity value obtained from the use of five features selected and step of similarity measurement. It showed that the similarity measurement step which adapts the working procedures of radiologist in examining the results of mammography screening was able to give better results. However, from 30 testing images of patients, four images were misclassified. They consist of two abnormal patient images which were classified as normal. If other images after passing through steps of image enhancement, feature extraction, and similarity measurement, the results of the classification is appropriate, but not with the images of these patients. After reviewing the features extraction result of the two images of these patients, it appears that there were some features of left and right images were not much different. There are a few characteristics of the patient’s breast image which were quite difficult to identify as abnormal medically. Mammogram results are not good when converted to digital images because the extracted features are less precise. This is due to the image of patients who have dense glands. Moreover, if in a feature graph seems that the similarity measurement of a normal patient is getting closer to zero and the value of similarity measurement of the abnormal patient is bigger than the similarity measurement value of normal patient, the feature will be a better option than other features when used as input classifier. It shows the difference value between normal and abnormal features makes classification easier. Then this feature can be used as input classifier.

A GUI was built after learning stage and testing stage was successfully done. The GUI was built by using push button to input left and right mammogram, push button to enhance the inputted images by median filter and CLAHE method, push button to extract the feature of inputted images, push button to measure the similarity measurement, and push button to classify the inputted images by backpropagation algorithm. Finally, the check box normal or abnormal was ticked according to the classification result.

Finally, the built system was simulated. First, the system was tried by inputting normal mammogram images and the result was shown in Figure 13. In Figure 13 normal check box was ticked. Then, the system was tried by inputting abnormal mammogram images and the result was shown in Figure 14. In Figure 14 abnormal check box was ticked. This shows that the method applied succeeded in classifying inputted mammogram correctly.

Table 3. Confusion matrix

Classification on 30 mammogram images		Actual values Positive (P)	Negative (N)
Classification Result	True	18	10
	False	0	2



```
Classifier output

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          28               93.3333 %
Incorrectly Classified Instances         2                6.6667 %
Kappa statistic                          0.8571
Mean absolute error                      0.0734
Root mean squared error                  0.216
Relative absolute error                 16.3826 %
Root relative squared error             45.7986 %
Total Number of Instances               30

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.000    1.000      0.900    0.947      0.866    0.975     0.990     AN
                 1.000    0.100    0.833      1.000    0.909      0.866    0.975     0.943     N
Weighted Avg.    0.933    0.033    0.944      0.933    0.935      0.866    0.975     0.974

=== Confusion Matrix ===

  a  b   <-- classified as
 18  2 |  a = AN
  0 10 |  b = N
```

Figure 11. Detailed accuracy by class

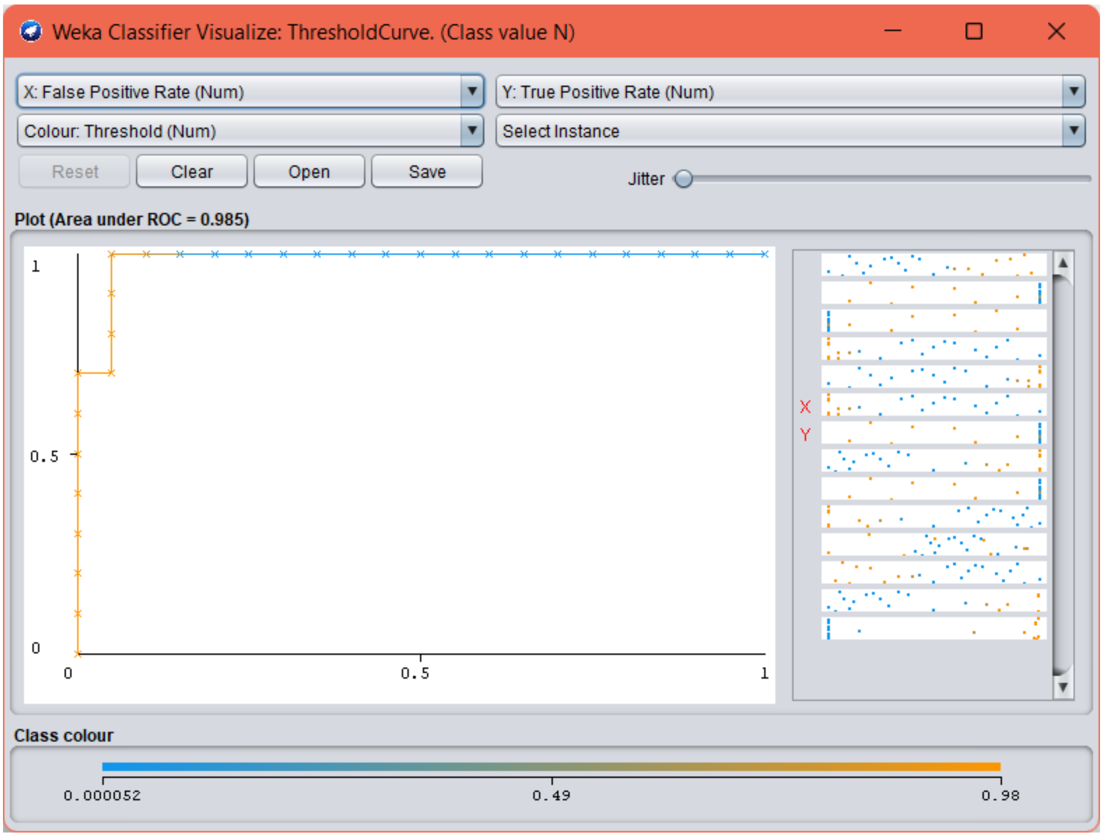


Figure 12. ROC curve of classification on 30 mammogram images

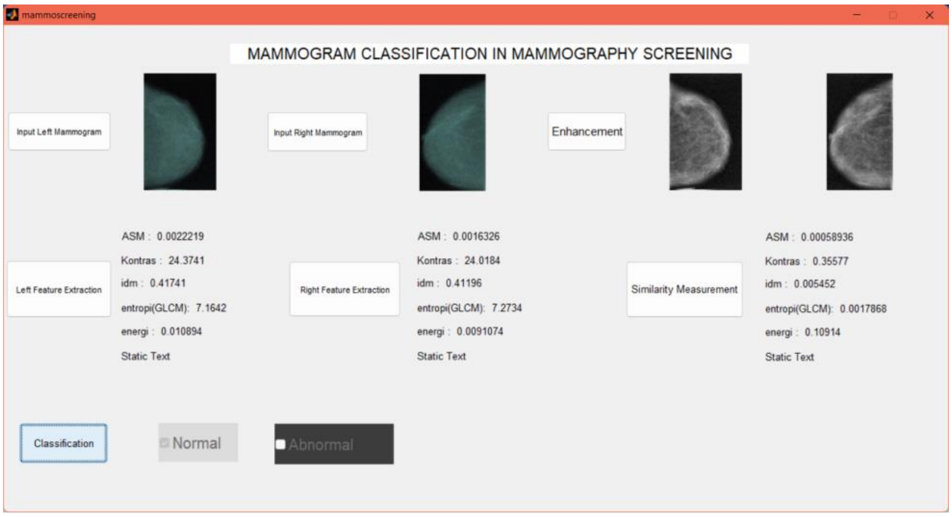


Figure 13. Normal checkbox ticked in MATLAB GUI of mammogram classification in mammography screening

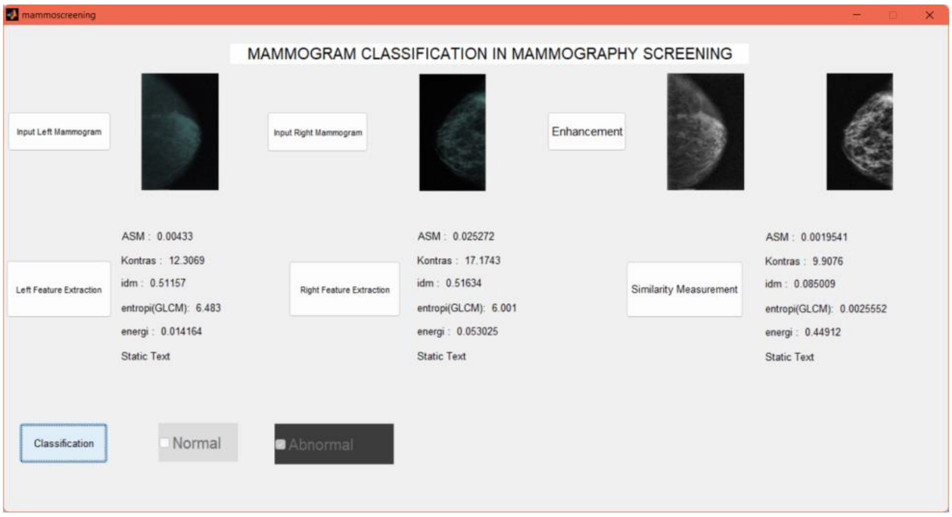


Figure 14. Abnormal checkbox ticked in MATLAB GUI of mammogram classification in mammography screening

4. CONCLUSION

Proposed similarity measurement which had performed before the classification can increase accuracy, sensitivity, and specificity in the classification result. The proposed method achieved accuracy of 0.933, sensitivity of 0.833, and specificity of 1.000. The high values of accuracy, sensitivity and high specificity value obtained from the use of five features selected and step of similarity measurement. Selected features used were ASM by GLCM, IDM by GLCM, contrast by GLCM, entropy by GLCM, and energy by histogram. Step of similarity measurement in this research based on symmetry measurement. The step was done by measuring the difference between the left and right mammogram of the patient. Previous research did CBIR-based similarity measurement. CBIR method did not adapt the working procedure of radiologists. It is only used one testing image so there was no symmetry measurement between left and right mammogram as radiologist did. Step of similarity measurement in this research adapted the working procedure of radiologists in examining screening mammography results that compared left mammogram with right mammogram to find if there is any asymmetry between them, then a radiologist will define which are normal and which are abnormal mammograms. Classification using similarity measurement can be used as a second opinion for radiologists. However, it requires improvement to obtain higher accuracy. It can be obtained by adding training data and improving image quality by more precise methods. Thus, there are no more cases of false negative.

REFERENCES

[1] International Agency for Research on Cancer of World Health Organization. “Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018.” uicc.org. https://www.uicc.org/news/global-cancer-data-globocan-2018. (accessed Nov. 09, 2020).

[2] D. H. Bhatia, “Breast Cancer in Asia,” *Gen Re (Series Risk Insights)*, vol. 5, pp. 1–7, 2016. [Online]. Available: https://media.genre.com/documents/ri16-4-en.pdf

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.

[4] The Global Cancer Observatory International Agency for Research on Cancer. “Cancer Incident in South-Eastern Asia International Agency for Research on Cancer.” Globocan. https://gco.iarc.fr/today/data/factsheets/populations/920-south-eastern-asia-fact-sheets.pdf. (accessed: May. 11, 2021).

[5] The Global Cancer Observatory International Agency for Research on Cancer. “Cancer Incident in Indonesia. International Agency for Research on Cancer.” Globocan. https://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf. (accessed: Nov. 11, 2021).

[6] S. Lei *et al.*, “Global Patterns of Breast Cancer Incidence and Mortality: A Population-Based Cancer Registry Data Analysis from 2000 to 2020,” *Cancer Communications*, vol. 41, no. 11, pp. 1183–1194, 2021, doi: 10.1002/cac2.12207.

[7] J. G. Elmore, “Chapter 1 - Breast Cancer Screening: Balancing Evidence With Culture, Politics, Money, and Media,” in *Breast Cancer Screening An Examination of Scientific Evidence*, 2016, pp. 1–27, doi: 10.1016/B978-0-12-802209-2.00001-2.

[8] K. H. Hwang *et al*., “Computer aided diagnosis (CAD) of breast mass on ultrasonography and scintimammography,” *Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005.*, 2005, pp. 187-189, doi: 10.1109/HEALTH.2005.1500435.

[9] C. -C. Jen and S. -S. Yu, “Automatic detection of abnormal mammograms in mammographic images,” *Expert System with Applications*, vol. 42, no. 6, pp. 3048–3055, 2015, doi: 10.1016/j.eswa.2014.11.061.

[10] C. Laronga, “Patient education: Breast cancer guide to diagnosis and treatment (Beyond the Basics).” up to date. https://www.uptodate.com/contents/breast-cancer-guide-to-diagnosis-and-treatment-beyond-the-basics. (accesed: Nov. 10, 2020).

[11] A. Aydiner, A. Igci, and A. Soran, *Breast Cancer: A Guide to Clinical Practice*, 1st ed. 20. Springer, 2019, doi: 10.1007/978-3-319-96947-3.

[12] A. Evans *et al.*, “Breast ultrasound: recommendations for information to women and referring physicians by the European Society of Breast Imaging,” *Insights Imaging*, vol. 9, pp. 449–461, 2018, doi: 10.1007/s13244-018-0636-z.

[13] A. Abubaker, “An Adaptive CAD System to Detect Microcalcification in Compressed Mammogram Images,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8, no. 6, pp. 133–138, 2017, doi: 10.14569/IJACSA.2017.080617.

[14] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Koward, and A. R. Feinstein, “Variability in Radiologist Interpretations of Mammograms,” *The New England Journal of Medicine*, vol. 331, no. 22, pp. 1493–1499, 1994, doi: 10.1056/NEJM199412013312206.

[15] P. J. Murray, G. Wivell, and E. Denton, “Breast cancer screening and diagnosis in the 21st century within the UK,” *Post Reproductive Health*, vol. 21, no. 3, pp. 105–111, 2015, doi: 10.1177/2053369115594954.

[16] K. K. Evans, T. M. Haygood, J. Cooper, A. -M. Culpan, and J. M. Wolfe, “A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast,” *Proc. Natl. Acad. Sci.* 2016, vol. 113, no. 37, doi: 10.1073/pnas.1606187113.

[17] H. A. Nugroho, N. Faisal, I. Soesanti and L. Choridah, “Analysis of digital mammograms for detection of breast cancer,” *2014 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 2014, pp. 25-29, doi: 10.1109/IC3INA.2014.7042595.

[18] V. P. Singh, A. Srivastava, D. Kulshreshtha, A. Chaudhary, and R. Srivastava, “Mammogram Classification Using Selected GLCM Features and Random Forest Classifier,” *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 14, no. 6, pp. 82–87, 2016. [Online]. Available: https://www.academia.edu/27067717/Mammogram_Classification_Using_Selected_GLCM_Features_and_Random_Forest_Classifier

[19] V. P. Singh, S. Srivastava, and R. Srivastava, “Automated and effective content-based image retrieval for digital mammography,” *Journal of X-Ray Science and Technology*, vol. 26, no. 1, pp. 29–49, 2018, doi: 10.3233/XST-17306

[20] A. S. Setiawan, Elysia, J. Wesley, and Y. Purnama, “Mammogram Classification using Law’s Texture Energy Measure and Neural Networks,” *Procedia Computer Science*, vol. 59, pp. 92–97, 2015, doi: 10.1016/j.procs.2015.07.341.

[21] J. Liu, J. Chen, Z. Huang, and X. Liu, “Design and Implementation of Content-based Medical Image Retrieval System on Mammograms,” *2011 4th International Conference on Biomedical Engineering and Informatics, BMEI*, 2011, pp. 237–240, doi: 10.1109/BMEI.2011.6098346.

[22] S. Molloi, H. Ding, and S. Feig, “Breast Density Evaluation Using Spectral Mammography, Radiologist Reader Assessment, and Segmentation Techniques: A Retrospective Study Based on Left and Right Breast Comparison,” *Academic Radiology*, vol. 22, no. 8, pp. 1–8, doi: 10.1016/j.acra.2015.03.017.

[23] H. Cho *et al.*, “Similarity evaluation in a content-based image retrieval (CBIR) CADx system for characterization of breast masses on ultrasound images,” *Medical Physics*, vol. 38, no. 4, 2011, doi: 10.1118/1.3560877.

[24] C. -H. Wei, S. Y. Chen, and X. Liu, “Mammogram retrieval on similar mass lesions,” *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 234–248, 2010, doi: 10.1016/j.cmpb.2010.09.002.

[25] D. T. Geddes, “Ultrasound imaging of the lactating breast: methodology and application,” *International Breastfeeding Journal*, vol. 4, no. 4, pp. 1–17, 2009, doi: 10.1186/1746-4358-4-4.

[26] K. M. Kanal *et al.*, “ACR – AAPM – SIIM Practice Guideline for Determinants of Image Quality in Digital Mammography,” *Journal of Digital Imaging*, vol. 26, pp. 10–25, 2013, doi: 10.1007/s10278-012-9521-3.

[27] N. Perry, M. Broeders, C. De Wolf, S. Törnberg, and L. V. Karsa, “European guidelines for quality assurance in breast cancer screening and diagnosis,” *Annals of Oncology*, vol. 19, no. 4, 2008, doi: 10.1093/annonc/mdm481.

[28] P. J. Lewis and N. J. McNulty, *Oxford American Handbook of Radiology*. USA: Oxford University Press, 2013. [Online]. Available: https://openlibrary.org/works/OL16532349W/Oxford_American_handbook_of_radiology

[29] V. S. Alfio, D. Costantino, and M. Pepe, “Influence of Image TIFF Format and JPEG Compression Level in the Accuracy of the 3D Model and Quality of the Orthophoto in UAV Photogrammetry,” *Journal of Imaging*, vol. 6, no. 5, 2020, doi: 10.3390/jimaging6050030.

[30] S. Punitha, A. Amuthan, and K. S. Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," *Future Computing and Infromatics Journal*, vol. 3, no. 2, pp. 348–358, 2018, doi: 10.1016/j.fcij.2018.10.005.
[31] E. Alimudin, H. A. Nugroho, and T. B. Adji, "Comparative Study on Mammogram Image Enhancement Methods According to the Determinant of Radiography Image Quality," in *Proceedings of Second International Conference on Electrical Systems, Technology and Information 2015 (ICESTI 2015)*, 2016, vol. 365, pp. 65–73, doi: 10.1007/978-981-287-988-2_6.
[32] E. Alimudin, H. A. Nugroho, and T. B. Adji, "Relevant Features for Classification of Digital Mammogram Images," *Proceedings of Second International Conference on Electrical Systems, Technology and Information 2015 (ICESTI 2015)*, 2016, pp. 115–122, doi: 10.1007/978-981-287-988-2_12.
[33] H. A. Nugroho, N. Faisal, I. Soesanti and L. Choridah, "Identification of malignant masses on digital mammogram images based on texture feature and correlation based feature selection," *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2014, pp. 1-6, doi: 10.1109/ICITEED.2014.7007907.
[34] M. M. R. Krishnan, S. Banerjee, and C. Chakraborty, C. Chakraborty, and A. K. Ray, "Statistical analysis of mammographic features and its classification using support vector machine," *Expert System with Application*, vol. 37, no. 1, pp. 470–478, 2010, doi: 10.1016/j.eswa.2009.05.045.
[35] T. Siriborvornratanakul, "An Automatic Road Distress Visual Inspection System Using an Onboard In-Car Camera," *Advances in Multimedia*, vol. 2018, 2018, doi: 10.1155/2018/2561953.
[36] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, 2020, doi: 10.1016/j.aci.2018.08.003.
[37] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC : A Family of Discriminant Measures for Performance Evaluation," *Australasian Joint Conference on Artificial Intelligence*, 2006, pp. 1015–1021, doi: 10.1007/11941439_114.

BIOGRAPHIES OF AUTHORS

Erna Alimudin is a lecturer at Cilacap State Polytechnic, Indonesia. She was a lectutrer at Sekolah Tinggi Teknologi Dumai since 2015 until 2018. She obtained Bachelor of Engineering Honours (S.T.) in Electrical Engineering from Universitas Hasanuddin, Indonesia. She also obtained her Master Degree's from Electrical Engineering from Universitas Gadjah Mada, Indonesia. Her research interests are image processing, decision support system, and embedded system. She was one of the recipients of research grants for novice lecturers by Ministry of Research and Technology/National Research and Innovation Agency of the Republic of Indonesia, in 2017 and 2019. She can be contacted at email: ernaalimudin@pnc.ac.id.

Hanung Adi Nugroho obtained Bachelor of Engineering Honours (S.T.) in Electrical Engineering from Universitas Gadjah Mada (UGM), Indonesia (2001), Master of Engineering (M.E.) in Biomedical Engineering from The University of Queensland, Australia (2005), Doctor of Philosophy (Ph.D.) in Image Processing from Universiti Teknologi PETRONAS, Malaysia (2012), Engineer (Ir.) in Electrical Engineering from Universitas Gadjah Mada, Indonesia (2018), and Professional engineer (IPM) in Electrical Engineering from Persatuan Insinyur Indonesia (PII), Indonesia (2018). He has been a lecturer in the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada (UGM) since 2002. Currently, he is an Associate Professor and also a Head of Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada (UGM). His research interests are signal and image processing and analysis, computer vision, medical imaging, medical instrumentation, statistical pattern recognition. He can be contacted at email: adinugroho@ugm.ac.id.

Teguh Bharata Adji is an expert in the field of natural language processing, artificial intelligence, image processing, decision support and group support system. He was an undergraduate and postgraduate degree at Electrical Engineering (S.T., M.T.), Universitas Gadjah Mada, Indonesia. He obtained his Master Degree's from Electrical Engineering (M.Eng.), Doshisha University, Japan. He also obtained a Doctorate Degree from Computer and Information Science (Ph.D), Universiti Teknologi Petronas, Malaysia. Now, he is an Associate Professor at the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada (UGM). His research interests are natural language processing, big data, content filter, machine translation, sentiment analysis and social media application. He can be contacted at email: adji.tba@gmail.com.