Feature Extraction Optimization with Combination 2D-Discrete Wavelet Transform and Gray Level Co-Occurrence Matrix for Classifying Normal and Abnormal Breast Tumors

Lulu M. Wisudawati[1], Sarifuddin Madenda[1], Eri P. Wibowo[1] & Arman A. Abdullah[2]

[1]Faculty of Computer Science and Information Technology, Gunadarma University, Indonesia

[2]Faculty of Medicine, Gunadarma University, Indonesia

Correspondence: Lulu M. Wisudawati, Faculty of Computer Science and Information Technology, Gunadarma University, Indonesia

Abstract

Breast cancer is one of the leading causes of death worldwide among women. According to GLOBOCAN Data, the International Agency for Research on Cancer (IARC), in 2012 there were 14.067.894 new cases of cancer and 8.201.575 deaths from cancer worldwide (Kementerian Kesehatan Republik Indonesia [KemenkesRI], 2015). Mammography is the most common and effective technique for detecting breast tumors. However, mammograms have poor image quality with low contrast. A Computer-Aided Detection (CAD) system has been developed to help radiologists effectively detect lesions on mammograms that indicate the presence of breast tumor. The feature extraction method in the CAD system is an important part of getting high accuracy results in classifying normal and abnormal breast tumors. By using the combination of 2D-Discrete wavelet transform and Gray-Level Co-Occurrence Matrix (GLCM) obtained an accuracy value of 100% on MIAS and UDIAT Database in classifying the presence of masses in the mammogram image and obtained an accuracy value of 93.8% for classifying normal, benign, and malignant. The proposed method has the potential to identify the presence of masses in the mammogram image as a decision support system to the radiologist.

Keywords: breast tumor, mammogram, CAD system, wavelet, GLCM, neural network

1. Introduction

Breast cancer is one of the leading causes of death worldwide among women today. According to GLOBOCAN data, the International Agency for Research on Cancer (IARC) in 2012 there were about 14.067.894 new cases of cancer and around 8.201.575 deaths due to cancer worldwide. Breast cancer is cancer with the highest percentage of new cases (43.3%) and the highest percentage of deaths (12.9%) in women in the world (Kementerian Kesehatan Republik Indonesia [KemenkesRI], 2015). Early detection of breast tumors can be done with medical imaging technology that is currently developing and it can be produced from various equipment in the medical field, such as Ultrasound (USG), MRI, CT-Scan / CAT-Scan, and Mammography (Fahnun, Mutiara, Wibowo, Arlan, and Latief, (2018).

Mammography is the most common and reliable method in the early detection of breast cancer. However, the high volume of mammograms to be read by physicians, the accuracy rate tends to decrease, and the automatic reading of digital mammograms becomes highly desirable (Velayutham and Thangavel, 2011). There are several types of abnormality in a mammogram, among them, microcalcifications and masses are the most common types. Masses can be described by three features, namely contour, boundary, and density. Computer-Aided Detection (CAD) system has been developed to aid radiologists effectively detect masses on mammograms that may indicate the presence of breast tumors. There are four stages in the CAD system for breast tumor classification: a) pre-processing b) detection and segmentation of suspected area c) post-processing, feature extraction (false positive reduction) d) evaluation, which is to evaluate the performance of CAD system. Figure 1 shows the overall process of the CAD system. The output of detected ROIs can be seen as a square. But it contains not only masses (blue square) but suspicious normal tissues (red square) as well. Therefore, false positive reduction is required to obtain only the real mass. In false positive reduction stage, the suspicious region that is normal but interpreted as mass are deleted (red dashed square) and the one interpreted as real masses are kept.
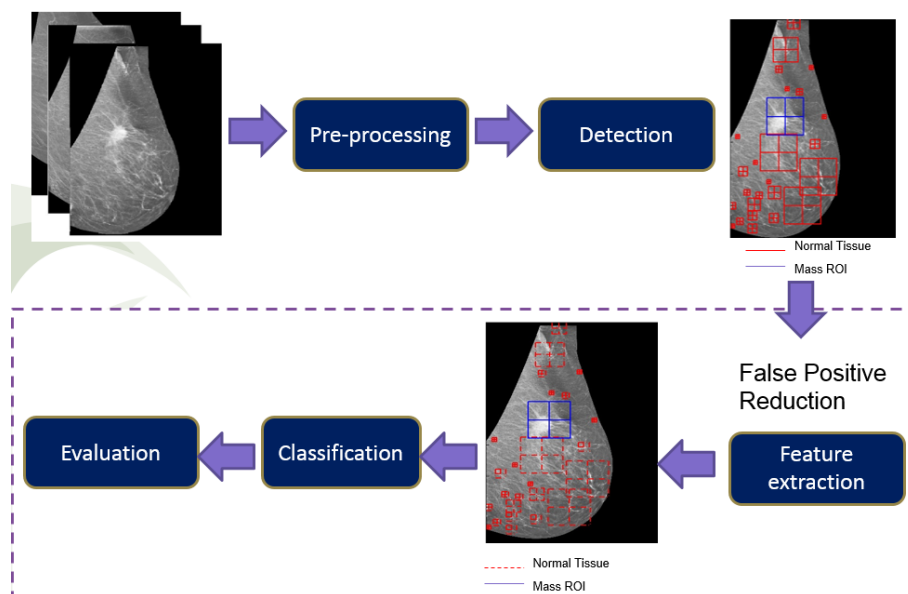
Figure 1. CAD System

In this study, our motivation is to develop a robust and discriminative feature extraction mechanism for false positive reduction (normal and abnormal classification) and normal, benign, and malignant classification to optimize the performance of CAD systems. The pre-processing and feature extraction process in the CAD system is an important stage in identifying the presence of tumors. A combination of 2D-Discrete Wavelet Transform and Gray Level Co-Occurrence Matrix (GLCM) methods is used to obtain optimal results. This paper is organized as follows: we present related works in Section 2. Our proposed method is described in Section 3. In Section 4, we present some experimental results to show the effectiveness of the proposed method. Finally, conclusions and limitations are given in Section 5.

2. Related Works

This section reviews the state-of-the-art about detecting the presence of the mass problem and pointing out their advantages and disadvantages. One of the most important stages in the development of the CAD system is a pre-processing process which aims to improve image quality such as contrast enhancement to obtain a better image visualization (Gandhamal, Talbar, Gajre, Hani, and Kumar, 2017), (Swaminathan and Gayathri, 2015), (Pawar and Talbar, 2018), (Mane and Kulhalli, 2015). Besides, there is a process of removing noise, one of them using the median filter which can maintain the edge information. Mohammed M. Abdelsamea and Bamatraf (2019) perform the pre-processing process by cropping unwanted areas.

Many techniques have been proposed to improve false positive reduction (normal and abnormal classification). False positive reduction depends on the Region of Interest (ROI) description. Various descriptors have been used to detect normal and masses areas in mammograms based on texture, gray-level, gradient and shape. A lot of research has been done on the textural analysis of mammographic images. Digital mammograms have specific characteristics, which are not all visual features can be used to describe relevant information. All classes for suspected tissue differ from shape, margins and tissue (Lestari, Madenda, and Massich, 2015). Milosevic, Jankovic, and Peulic (2014) propose an approach using GLCM in the process of feature extraction with 20 texture descriptors. This method achieved an accuracy of 83.7% with sensitivity and specificity of 80.7% and 86.7%, respectively. The experiments were conducted on local database with 300 images.

Pratiwi, Alexander, Harefa, and Nanda (2015) approach Radial Basis Function Neural Network (RBFNN) for mammograms classification based on Gray Level Co-occurrence Matrix (GLCM) texture-based features. They extract the features in the suspected areas of being mass. The computational experiments show that RBFNN is better than Back-propagation Neural Network (BPNN) in performing breast cancer classification. The result using RBFNN achieved an accuracy of 93.98% on MIAS database, which is 14% higher than BPNN and the accuracy of benign and malignant classification is 94.29% which is 2% higher than BPNN.

Biswas, Nath, and Roy (2016) perform an automated CAD system to classify the breast tissues as normal and abnormal. Artifacts are removed using ROI extraction process and noise has been removed by the 2D median filter. Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm is used to improve the appearance of the image. They perform normalization by resizing the image to 128 x 128 pixels as well. The texture features are extracted using Gray Level Co-Occurrence Matrix (GLCM) of the region of interest (ROI) of a mammogram. The proposed method with 3NN classifier on MIAS database by giving 95% accuracy, 100% sensitivity and 90% specificity to classify mammogram images as normal or abnormal. Ergin, Esener, and Yuksel (2016) used texture descriptor on mammogram images with 88-dimensional features and classification processes using the Fisher's Linear Discrimination Analysis (FLDA) method and the accuracy obtained 72.93%.

In addition to texture features, multiresolution analysis is used effectively in image features for classifying normal and abnormal breast tumors. Mammogram image is decomposed into several sub-images that preserve information about both low and high frequencies. Many researchers worked on a multiresolution analysis of mammograms based on wavelets transform by using different types of wavelet family and feature spaces. Ucar and Kocer (2017) used Wavelet Neural Network (WNN) and classical Neural Network in normal and abnormal classification. The best estimated result of WNN is 98.90%. The best result for various artificial neural networks is calculated as 95.49%. Putra (2018) proposed Local Binary Pattern to all the detailed coefficients from 2D-DWT of the region of interest (ROI) of a mammogram to generate a feature matrix. The performance of the proposed scheme using Neural Network classifier can produce high accuracy that is 92.71% on MIAS and DDSM database. Pawar and Talbar (2018) increased the performance of Computer-Aided Diagnosis (CAD) by reducing false positives (FP). FP reduction consists of feature mining from the ROIs using proposed local sparse curvelet coefficients followed by classification using an artificial neural network (ANN). The performance of the proposed algorithm has been validated using the local datasets as TMCH (Tata Memorial Cancer Hospital) and achieved an accuracy of 98.3%.

Referring to the advantages and disadvantages of the methods and algorithms that have been developed by previous researchers, in this paper, we propose an effective feature extraction algorithm using a combination of two-dimensional discrete wavelet transform (2D-DWT) based multi-resolution analysis and compute texture feature using GLCM. A Supervised classifier using Backpropagation Neural Network is used to classify normal and abnormal breast tumors.

3. Method

The first process in diagnosing breast tumors is to identify the presence of tumors or classify normal (no tumor) and abnormal (indicated the presence of tumor). There are several stages in classifying normal and abnormal mammogram images: image acquisition, preprocessing, feature extraction, and classification.
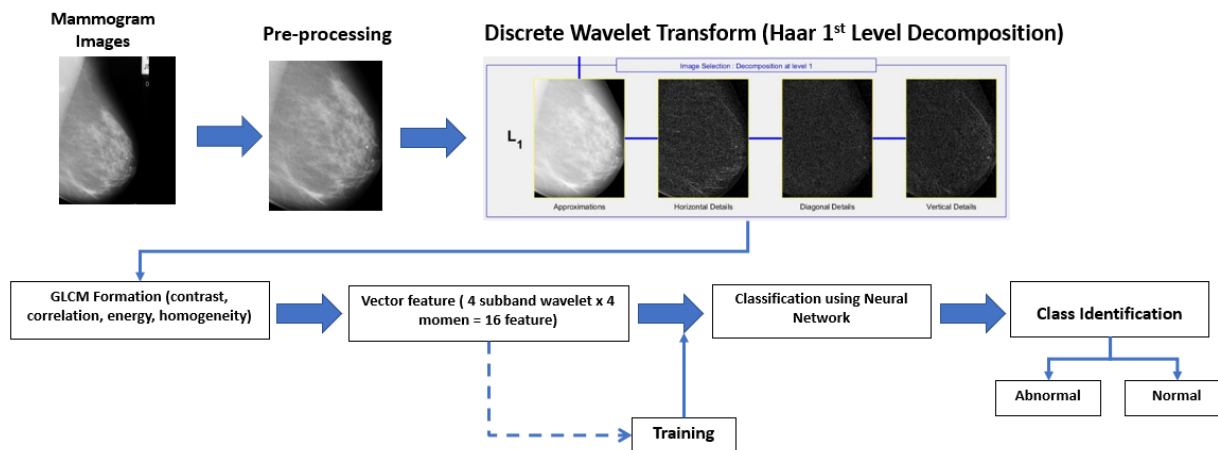

Figure 2. Overview of the Proposed Method

The initial stage in diagnosing breast tumors (see Figure 2) is image acquisition which is a mammogram image obtained from MIAS and UDIAT (Hospital of Sabaddel). Pre-processing is performed by removing the irrelevant areas of the mammogram image and extract the suspected areas (ROI). Feature extraction is conducted to obtain the characteristics of the mammogram image. Since digital mammograms have specific characteristics, not all visual features can be used to correctly describe the relevant information. Extracted ROI is decomposed into a 1-level discrete wavelet transform using Haar wavelets preserving the high and low frequency information. It

generates four images: approximation (LL), horizontal detail coefficients (LH), vertical detail coefficients (HL), and diagonal detail coefficients (HH). The feature extraction process is performed using GLCM and analysis of texture pattern measurements of each ROI using four statistical texture descriptors: contrast, correlation, energy, and homogeneity. The length of vector features is 16 features (4 sub-bands wavelet and 4 moments). Backpropagation Artificial Neural Network method is used to identify the presence of masses on mammogram images.

3.1 Pre-Processing

The pre-processing are fundamental steps in the medical image processing to produce better image quality for segmentation and feature extractions. Mammograms have low contrast, background noise, artifacts (see Figure 3). Therefore, pre-processing is needed to improve image quality, remove unwanted noise, preserves the edges within an image, and enhance the image. There are different types of filtering techniques in pre-processing, in this study using median filter operations, thresholding, and mathematical morphological operations (see Table 1).
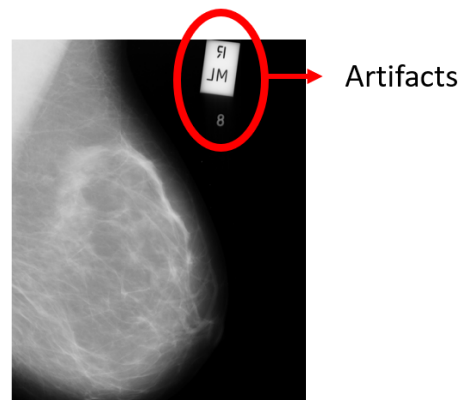

Figure 3. Mammogram Artifact

Table 1. Process of Removing Noise and Artifacts

Mammogram Image	Median Filter	Thresholding	Morphological Operation	Result

After removing the label, the extraction of the region of interest (ROI) is carried out automatically and adaptively. Description of ROI is an area of breast tissue (mammae area). The automatic cropping process is conducted by the proposed method and adaptively adjusts to the mammogram characteristics. The proposed method to extract ROI for the left orientation mammogram image can be seen in Algorithm 1:

Algorithm 1. Automatic Image Cropping (Left Orientation)

Input: The result of removing noise and artifacts (Img)

Output: Region of Interest (mammae area)

1. Start

2. Read input image ($Img_{n,m}$)

3. Define the start point (x_0,y_0) and endpoint (x_1,y_1) x_0=0 ; y_0=0 and x_1=0 ; y_1=0

4. Read each point in the input image from the right position (horizontal) and get the new position x_0 and y_0 at the point where the intensity found (the right boundary of the intensity between tissue and background):

   If $I_{(n,m)} > 0$ and $x_0 = 0$ Then x_0= m and y_0=n

5. Read each point in the input image from the bottom position (vertical), and get the new position x_1 and y_1 at the point where the intensity found (the bottom boundary of the intensity between tissue and background):

   If $I_{(n,m)} > 0$ and $x_1 = 0$ Then x_1= m and y_1=n

6. Get the mammae area (region of interest): 2 times the length and width of the x and y positions (Icrop)

7. End

In Algorithm 1, the process begins by reading the input image which is left-oriented mammogram image. Defined the start point (x_0, y_0) = 0 and endpoint (x_1, y_1) = 0. Search for the appearance of the intensity value starts from the right position of the image horizontally and get a new position x_0 and y. Then search for the appearance of the intensity value from the bottom position of the image vertically and get a new position x_1 and y_1. Cropping mammae area is conducted by calculating 2 times the length and width of the x and y positions. Algorithm 2 describes the image cropping algorithm mammogram for right orientation which is carried out the same process in Algorithm 1, but the difference is reading the image point starting from the left position of the image.

Algorithm 2. Automatic Image Cropping (Right Orientation)

Input: The results of removing noise and artifacts (Img)

Output: Region of Interest (mammae area)

1. Start

2. Read input image ($Img_{n,m}$)

3. Define the start point (x_0,y_0) and endpoint (x_1,y_1) x_0=0 ; y_0=0 and x_1=0 ; y_1=0

4. Read each point in the input image from the left position (horizontal), and get the new position x_0 and y_0 at the point where the intensity found (the right boundary of the intensity between tissue and background):

   If $I_{(n,m)} > 0$ and $x_0 = 0$ Then x_0= m and y_0=n

5. Read each point in the input image from the bottom position (vertical), and get the new position x_1 and y_1 at the point where the intensity found (the bottom boundary of the intensity between tissue and background):

   If $I_{(n,m)} > 0$ and $x_1 = 0$ Then x_1= m and y_1=n

6. Get the mammae area (region of interest): 2 times the length and width of the x and y positions (Icrop)

7. End

The illustration of the Algorithm 1 can be seen in Figure 4. where A is a mammogram image (preprocessing results) with left orientation, B describes direction from the right and bottom to find the appearance of intensity points. C describes how to get area mammae: 2 times the length and width between x and y points and D describes the extracted mammae area.
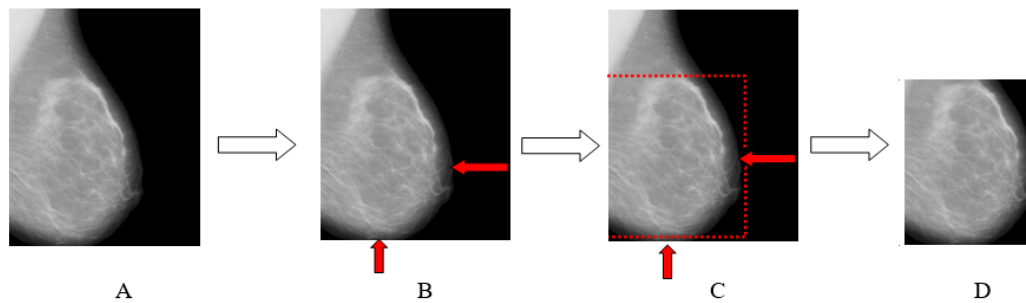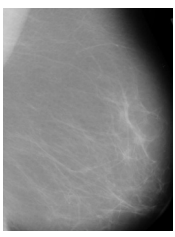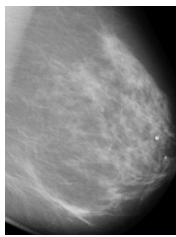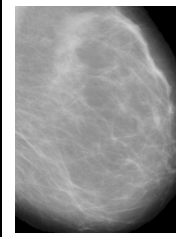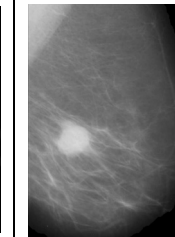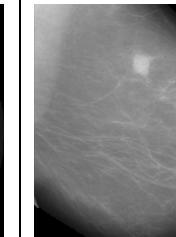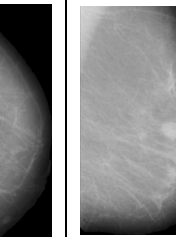
Figure 4. Illustration of ROI Extraction Algorithm (Example: Left Orientation)

The results of ROI extraction can be seen in Table 2, these results indicate that the mammae area (normal and abnormal) has been obtained and the ROI is used as input in the feature extraction process.

Table 2. The Results of ROI Extraction (Normal and Abnormal Mammae Area using MIAS Database)

3.2. Feature Extraction

Feature extraction is an important step in image classification. Feature extraction aims to extract proper features to distinguish different textures micropatterns. It has been proven to help differentiate mass and normal tissue as well as benign and malignant masses. The proposed feature extraction method is a combination of Wavelet and GLCM and it can be seen in Figure 5. Discrete wavelet transform (DWT) decomposes the ROI into 4 orthogonal sub-band and performed GLCM then compute four statistical features of texture (contrast, correlation, energy, and homogeneity).
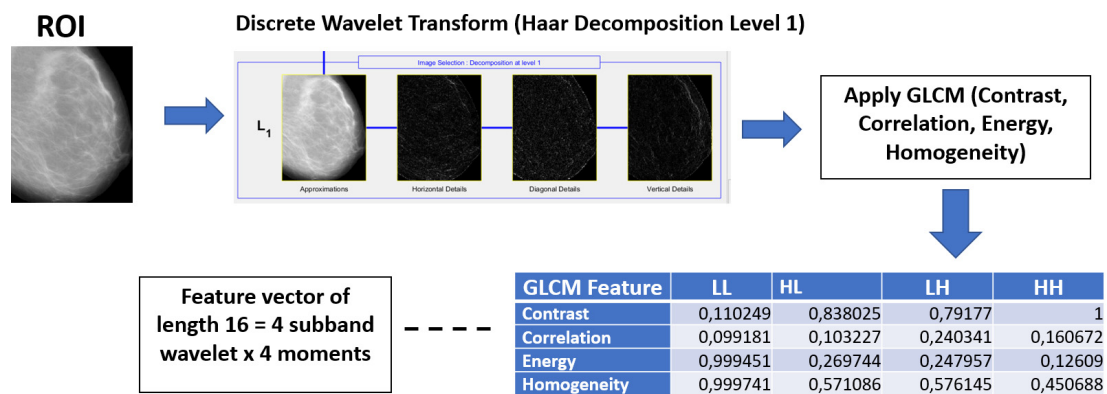

GLCM Feature	LL	HL	LH	HH
Contrast	0,110249	0,838025	0,79177	1
Correlation	0,099181	0,103227	0,240341	0,160672
Energy	0,999451	0,269744	0,247957	0,12609
Homogeneity	0,999741	0,571086	0,576145	0,450688

Figure 5. Feature Extraction Process

3.2.1 Discrete Wavelet Transform (DWT)

Discrete wavelet transform (DWT) decomposes the image into 4 orthogonal sub-band: low-low (LL), high low (HL), low high (LH), and high-high (HH) consisting of approximation, horizontal, vertical, and diagonal information. The implementation of discrete wavelet transforms can be done by passing high frequency and low frequency signals. Extracted ROI is decomposed into a 1-level discrete wavelet transform using Haar wavelets. The results of wavelet decomposition can be seen in Figure 6. The approximations image is the smoothed

version of the original image and it contains global information that is similar to the original image with the number of rows and the number of columns being half of the original image. Horizontal, vertical, and diagonal contain the detail and represent the fluctuations of the pixel intensity in horizontal, vertical, and diagonal directions and they have low-intensity areas, whereas areas with high intensity are only found on the edges of the image object, therefore the shape pattern of mammae area is obtained.
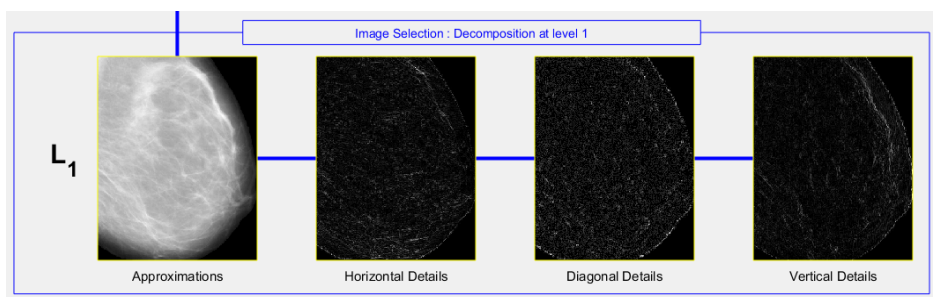

Figure 6. The Results of Wavelet Decomposition

(Approximation, Horizontal Details, Diagonal Details and Vertical Details)

3.2.2 GLCM Formation

The next step is the formation of GLCM. The stage of GLCM formation using orientation 0^0 and distance d = 1 which means the coordinates (x, y) is [0,1]. After determining the direction, specify the number of gray-levels `graycomatrix` uses to scale the image by using the 'NumLevels' parameter, and the way that `graycomatrix` scales the values using the 'GrayLimits' parameter. In this study, graycomatrix using Numlevel is 32, which means 2^5 or 5 bits and using graylimit minimum and maximum grayscale values in the input image as limits.

Feature extraction with texture analysis is conducted by taking the characteristics of grayscale image to distinguish one object from another object. The object is extracted based on statistical measurements namely contrast, energy, correlation and homogeneity. Each image has a feature vector of length 16 (4 sub-bands wavelet x 4 statistical measures). Contrast measures the local contrast of an image. Correlation provides a correlation between the two pixels in the pixel pairs. Energy measures the number of repeated pairs. The energy is expected to be high if the occurrence of repeated pixel pairs is high. Homogeneity measures the local homogeneity of a pixel pair. The homogeneity is expected to be large if the gray-levels of each pixel's pair are similar.

3.3 Backpropagation Artificial Neural Network (BP-ANN)

ANN is used as a classifier. We choose ANN because of its capability to learn from examples and capture the functional relationships among the hard description of data. Figure 7 show ANN classifier model and it has a two-layers feed-forward backpropagation network with sigmoid transfer functions. The backpropagation is based on Levenberg-Marquardt. We train the network several times with different amounts of hidden layers and neurons. The best results are obtained with 10 neurons. The feature vector obtained is used as input in the BP-ANN training.

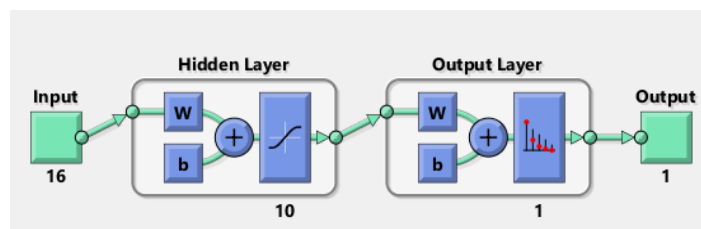Figure 7. Backpropagation Artificial Neural Network

4. Results and Discussion

4.1 Dataset

The database used in this study are mammogram images from MIAS (Mammographic Image Analysis Society) database: 57 normal and 120 abnormal images with a size of 1024 pixels x 1024 pixels and database obtained from UDIAT (Sabaddell Hospital): 52 abnormal images (19 benign and 33 malignant) (Tortajada et al., 2014).

4.2 Feature Extraction Results using Haar Wavelet and GLCM

The results of feature extraction using wavelet and GLCM for normal and abnormal classification (see Table 3) indicate that the contrast values in normal images in the LL sub-band are lower than the contrast values in abnormal images. That means the gray-levels of each pixel pair are similar. Whereas the LH, HL, and HH sub-band in the abnormal image shows that the energy and homogeneity values have a greater value than the normal image. This describes that the LH, HL and HH sub-band have high repeated pixel pairs and the gray-levels pixel pairs are similar. Targets or labels for normal and abnormal images are marked with a value of 1 and 0, respectively.

Table 3. The Results of Feature Extraction for Normal and Abnormal Classification

Wavelet Haar	GLCM Feature	Normal			Abnormal		
LL	Contrast	0.110249	0.82311	1.042705	2.100875	2.039434	2.101577
	Correlation	0.099181	0.90934	0.075774	0.99047	0.991463	0.992343
	Energy	0.999451	0.988661	0.995293	0.177657	0.069975	0.166695
	Homogeneity	0.999741	0.998486	0.997833	0.780077	0.735524	0.791733
LH	Contrast	0.79177	0.760014	0.835127	0.871523	0.742436	0.854165
	Correlation	0.240341	0.171935	0.298689	0.191402	0.226382	0.212118
	Energy	0.247957	0.279087	0.183297	0.307806	0.601763	0.372826
	Homogeneity	0.576145	0.589415	0.537664	0.777693	0.894719	0.809177
HL	Contrast	0.838025	0.856362	0.89576	1	1	1
	Correlation	0.103227	0.062627	0.125204	0.125696	0.091209	0.127587
	Energy	0.269744	0.255011	0.230984	0.310548	0.619217	0.37846
	Homogeneity	0.571086	0.551938	0.542024	0.77733	0.896893	0.809317
HH	Contrast	1	1	1	0.640986	0.408427	0.616771
	Correlation	0.160672	0.094231	0.200905	0.130385	0.092463	0.134886
	Energy	0.12609	0.139591	0.102557	0.381405	0.761287	0.457826
	Homogeneity	0.450688	0.449253	0.434299	0.810965	0.938519	0.842251
Target/label		1	1	1	0	0	0

4.3 Classification Results using Artificial Neural Network

Classification results using Artificial Neural Network are carried out by cross-validation which consists of 70% data training, 15% data validation, and 15% data testing. Confusion matrix in Figure 8 shows an accuracy value of 100% using MIAS and UDIAT database (it shows that 100% of the system can classify tumors correctly) as well as sensitivity and specificity of 100% (representing 100% normal and abnormal images that are correctly classified by the system). The test results show that 56.3% of data testing or 9 abnormal images can be correctly diagnosed by the system and 43.8% of data testing or 7 normal images can be correctly diagnosed by the system. Normal and abnormal classification processes are also performed using the MIAS database and get an accuracy value of 100% and a sensitivity and specificity value of 100%.

Figure 8. Confusion Matrix Normal and Abnormal Classification

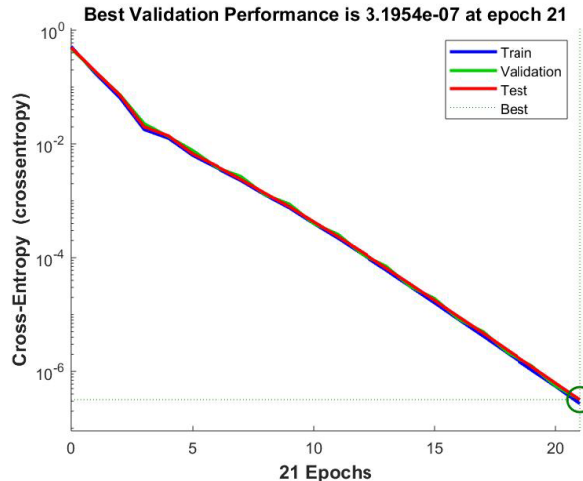Training performance an error goal (MSE) of 3.1954e-07 was achieved at epoch 21 as shown in Figure 9.

Figure 9. Best Validation Performance

This study conducted experiments in classifying normal, benign, and malignant tumors as well. The results of the confusion matrix can be seen in Figure 10 which consists of 16 data testing (normal, benign, and malignant). This result shows that 43.8% of testing data or 7 normal images can be correctly classified by the system. 12.5% of testing data or 2 benign images can be correctly classified by the system. 37.5% of testing data or 6 malignant images can be correctly classified by the system. The testing result shows that there is a false positive 6.3% of testing data or 1 normal image is classified by the system as a benign. Therefore, the accuracy obtained is 93.8%.

Figure 10. Confusion Matrix Normal, Benign, and Malignant

4.4 Comparison of Normal and Abnormal Classifications Performance

In this section, we present the comparison of our proposed method with various method present in the state of the art. However, it is hard to critically compare due to the evaluation of the methods conducted by different databases. Although the method performed using the same database but in the sample mammogram selection process are not the same. The different total number of mammograms used in different research works also one of the reasons. Furthermore, a different experimental setup is used for example k-nfold validation with varying k-value, classifier method is not the same for all the methods as well as preprocessing and segmentation process can be effect for classification results. However, our aim is to have general trends of performance comparison and we compare our method with other techniques on the basis of accuracy in Table 4 and Table 5.

Table 4. Comparison of Normal and Abnormal Performance (Accuracy, Sensitivity (*Sn*), Specificity (*Sp*))

Author	Database	Feature Extraction	Classifier	Accuracy	Sn	Sp
(Milosevic et al., 2014)	Local Database	GLCM	SVM	83.7%	80.7%	86.7%
(Milosevic et al., 2014)	MIAS	GLCM	SVM	62%	20.4%	87.2%
(Pratiwi et al., 2015)	MIAS	GLCM	RBFNN	93.98%	-	-
(Biswas et al. ,2016)	MIAS	GLCM	3NN	95%	100%	90%
(Ergin et al., 2016)	MIAS	GLCM	FLDA	72.39%	-	-
(Ucar and Kocer, 2017)	Local Database	Wavelet	ANN	95.49%	-	-
(Putra, 2018)	MIAS, DDSM	2D DWT+ LBP	ANN	92.1%	91%	94%
(Pawar and Talbar, 2018)	TMCH	LBP	ANN	98.30%	-	-
Proposed Method	MIAS	Wavelet+ GLCM	ANN	100%	100%	100%
Proposed Method	MIAS+UDIAT	Wavelet+ GLCM	ANN	100%	100%	100%

Firstly, we discuss the comparison dealing with normal and abnormal classification. Table 4. is a comparison of normal and abnormal performance from some researchers. Performance results are measured by accuracy, sensitivity *(Sn)* and specificity (*Sp*). The works of Milosevic et al. (2014) obtained an accuracy value of 83.7%, sensitivity value (*Sn*) of 80.7% and specificity value of (*Sp*) 86.7% using local database and accuracy value of 62% using MIAS database. Pratiwi et al. (2015) obtained an accuracy value of 93.98% with RBFNN classifier and Biswas et al. (2016) obtained an accuracy value of 95% with a sensitivity value of 100% and a specificity of 90% using 3NN classifier. Ergin et al. (2016) obtained an accuracy value of 72.39% using 322 database FLDA classifier. In addition, Ucar and Kocer (2017) used Wavelet for feature extraction method and ANN classifier obtained an accuracy value of 95.49%. Putra (2018) using 2D DWT+ LBP obtained accuracy values of 92.1% with sensitivity value (Sn) of 91% and Specificity value of (Sp) 94%. Pawar and Talbar (2018) using LBP

obtained an accuracy of 98.3%. Our proposed method using combination wavelet and GLCM for MIAS and UDIAT database obtained an accuracy of 100%. This result shows an increase in performance from previous researchers that used GLCM and wavelet method as feature extraction.

Table 5. Comparison of Normal, Benign and Malignant Performance

Experiment	Author	Database	Feature Extraction	Classifier	ROI	Accuracy
Normal, Benign, Malignant	(Elizabeth et al., 2016)	MIAS	MCEEMDAN	NN	317	96.7%
	Proposed Method	**MIAS-UDIAT**	**Wavelet-GLCM**	**NN**	**107**	**93.8%**

Table 5. shows the comparison of normal, benign, and malignant performance. The previous researcher, Elizabeth et al. (2016) used Multidimensional Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (MCEEMDAN) as a feature extraction method on MIAS database (317 images) obtained an accuracy of 96.7% for normal, benign, malignant classification. The proposed method obtained an accuracy of 93.8% using 2D-DWT and GLCM. However, the number of databases used is different so that it can affect the accuracy value.

5. Conclusion

The Computer-Aided Diagnosis (CAD) system was successfully developed in classifying mammogram images into normal and abnormal breast tumors and classifying normal, benign, and malignant. By using a combination of 2D-Discrete Wavelet Transform and Gray Level Co-Occurrence Matrix (GLCM) obtained an accuracy value of 100% in identifying the presence of masses in the mammogram image and obtained an accuracy for classifying normal, benign and malignant of 93.8%.

The feature extraction stage is an important step in the process of classifying breast tumors. A large number of features are obtained in sub-vector resolution by performing wavelet decomposition in 4 sub-band (LL-HL-LHHH) and four GLCM statistical measurements (contrast, energy, correlation, homogeneity). Therefore, the number of vector features is large and most likely many features are redundant. To overcome this problem, the next research will be carried out using a database with a larger number and perform a feature selection method to eliminate irrelevant features, reduce the dimensions of vector features and improve accuracy.

Acknowledgment

Competing Interest

The authors declare that they have no competing interests

References

Abdelsamea, M. M., Mohamed, M. H. & Bamatraf, M. (2019). Automated classification of malignant and benign breast cancer lesions using neural networks on digitized mammograms. *Cancer Informatics, 18*, 1–3. Sage Journal. https://doi.org/10.1177/1176935119857570PMid:31244522 PMCid:PMC6580711

Biswas, R., Nath, A. & Roy, S. (2016). *Mammogram classification using gray-level co-occurrence matrix for diagnosis of breast cancer,* 161– 166. https://doi.org/10.1109/ICMETE.2016.85

Caroline, B. Elizabeth & N. Vaijayanthi. (2016). Feature Extraction of Digital Mammogram Based on Multidimensional Complete Ensemble Empirical Mode Decomposition with Adaptive Noise". *International Journal of Advance Engineering Technology*

C Velayutham & K Thangavel. (2011). A novel feature extraction method using spectral shape in digital mammogram image. *In Information and Communication Technologies (WICT)*, 2011 World Congress on, pages 835840. IEEE, 2011. https://doi.org/10.1109/WICT.2011.6141356

Ergin, S., Esener, I. & Yuksel, T. (2016). A genuine glcm-based feature extraction for breast tissue classification on mammograms. *International Journal of Intelligent Systems and Applications in Engineering*, 124– 124. https://doi.org/10.18201/ijisae.269453

Fahnun, B. U., Mutiara, A. B., Wibowo, E. P., Arlan, J. & Latief, A. (2018). Filtering techniques for noise reduction in liver ultrasound images. *phdthesis, Program Doktor Teknologi Informasi Universitas Gunadarma*, 261–266. https://doi.org/10.1109/EIConCIT.2018.8878547

Gandhamal, A. P., Talbar, S., Gajre, S., Hani, A. F. M. & Kumar, D. (2017). Local gray level s-curve transformation a generalized contrast enhancement technique for medical images. *Computers in Biology and Medicine*, 83. https://doi.org/10.1016/j.compbiomed.2017.03.001 PMid:28279861

Kementerian Kesehatan Republik Indonesia [KemenkesRI] (2015). *Pusat data dan informasi kementerian republik indonesia.* Infodatin.

Lestari, D. & Madenda, S. & Massich, J. (2015). A Segmentation Algorithm for Breast Lesion Based on Active Contour Model and Morphological Operations. *Advanced Science, Engineering and Medicine, 7*, 920-924. 10.1166/asem.2015.1786. https://doi.org/10.1166/asem.2015.1786

Listia, R. & Harjoko, A. (2014). Klasifikasi massa pada citra mammogram berdasarkan gray level cooccurence matrix (glcm). *IJCCS, 8*(1), 59–69. ISSN: 1978-1520.

Mane, S. A. & Kulhalli, D. K. V. (2015). Mammogram image features extraction and classification for breast cancer detection. *International Research Journal of Engineering and Technology (IRJET), 02*(07).

Milosevic, M., Jankovic, D. & Peulic, A. (2014). Comparative analysis of breast cancer detection in mammograms and thermograms. *Biomedizinische Technik*, 60. https://doi.org/10.1515/bmt-2014-0047 PMid:25720034

Pawar, M. & Talbar, S. (2018). Local entropy maximization-based image fusion for contrast enhancement of mammogram. *Journal of King Saud University Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2018.02.008

Pawar, M., Talbar, S. & Dudhane, A. (2018). Local binary patterns descriptor based on sparse curvelet coefficients for false-positive reduction in mammograms. *Journal of healthcare engineering*. https://doi.org/10.1155/2018/5940436 PMid:30356422 PMCid:PMC6178513

Pratiwi, M., Alexander, Harefa, J. & Nanda, S. (2015). *Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network.* https://doi.org/10.1016/j.procs.2015.07.340

Putra, A. J. (2018). Mammogram classification scheme using 2d discrete wavelet and local binary pattern for detection of breast cancer. *Journal of Physics: Conference Series*, https://doi.org/10.1088/1742-6596/1008/1/012004

Swaminathan, A. & Gayathri, S. (2015). Mammogram image enhancement by two-stage adaptive histogram equalization. *Optik – International Journal for Light and Electron Optics*, 126. https://doi.org/10.1016/j.ijleo.2015.07.069

Tortajada, M., Oliver, A., Marti, R., Ganau, S., tortajada gimenez, L., Sentis, M., Freixenet, J. & Zwiggelaar, R. (2014). Breast peripheral area correction in digital mammograms. *Computers in Biology and Medicine, 50*. https://doi.org/10.1016/j.compbiomed.2014.03.010 PMid:24845018

Ucar, K. & Kocer, H. E. (2017). Breast cancer classification with wavelet neural network. *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1–5. https://doi.org/10.1109/IDAP.2017.8090347