❏   60

Feature Selection Mammogram based on Breast Cancer Mining

Shofwatul 'Uyun[1], Lina Choridah[2]
[1]Department of Informatics, Faculty of Science and Technology, Sunan Kalijaga State Islamic University, Yogyakarta, Indonesia
[2]Department of Radiology, Faculty of Medicine, Gadjah Mada University, Yogyakarta, Indonesia

Article Info	ABSTRACT
	The very dense breast of mammogram image makes the Radiologists often have difficulties in interpreting the mammography objectively and accurately. One of the key success factors of computer-aided diagnosis (CADx) system is the use of the right features. Therefore, this research emphasizes on the feature selection process by performing the data mining on the results of mammogram image feature extraction. There are two algorithms used to perform the mining, the decision tree and the rule induction. Furthermore, the selected features produced by the algorithms are tested using classification algorithms: k-nearest neighbors, decision tree, and naive bayesian with the scheme of 10-fold cross validation using stratified sampling way. There are five descriptors that are the best features and have contributed in determining the classification of benign and malignant lesions as follows: slice, integrated density, area fraction, model gray value, and center of mass. The best classification results based on the five features are generated by the decision tree algorithm with accuracy, sensitivity, specificity, FPR, and TPR of 93.18%; 87.5%; 3.89%; 6.33% and 92.11% respectively.

Corresponding Author:

Shofwatul 'Uyun,
Department of Informatics,
Faculty of Science and Technology,
Sunan Kalijaga State Islamic University,
Jl. Marsda Adisucipto No. 1 Yogyakarta, 55281, Indonesia.
Email: Shofwatul.uyun@uin-suka.ac.id

1.   INTRODUCTION

Breast cancer and cervical cancer are types of cancer that cause the highest mortality in women in Indonesia. Based on data [1] there are 330,000 cancer patients in Indonesia and the highest cancer prevalence is found in Yogyakarta Special Region of 4.1%. Breast cancer has established risks (e.g. family history, obesity, having dense breast) and emerging risks (e.g. low of vitamin D levels, unhealthy life style); therefore, the early detection can be conducted to reduce the mortality of breast cancer. In fact, if cancer are found in the early stage, there will be a great cure rate. However, mostly the cases of breast cancer in Indonesia are found in the advance stage because of low awareness. Mammography is one of the imaging technologies that can be used both for screening and for the diagnosis of breast cancer. Based on the BI-RADS lexicons for Mammography 2013, a hyperdensity mass with an irregular shape and spikulated margin is associated with malignancy. Other suspicious morphology is amorphous, coarse heterogenous, fine pleomorphic and fine linear or fine-linear branching calcification [2].

Currently a computer-aided system using the mammogram image with several different purposes has been developed, among others, to determine the level of risk of breast cancer [3], to detect the location considered abnormal in the image mammogram that is commonly called the CADE system [4], and to

diagnose the type of breast cancer considered as RoI on the mammogram image that is commonly called the CADx system. The use CADx is as the second opinion in diagnosing the breast cancer based on the reading of the mammogram image.

In general, there are several stages to develop a computer-aided diagnosis system (CADx), among others: image acquisition, pretreatment, feature extraction, feature selection, classification and testing. At each stage, it needs to do the right choice of algorithm in order to be able to provide the diagnosis result accurately. In principle, the work pattern of CADx system follows the work pattern of pattern recognition system. One important factor that determines the success or failure in pattern recognition system is the use of the right features. According to [5] the right feature selection is a critical stage because the right features makes the pattern recognition system capable to distinguish between one object from another one in accordance with the characteristics of the object, one based on improved document frequency for the text classification [6]. Therefore, it is necessary to do the feature selection on a mammogram that is able to distinguish between benign from malignant lesions on the mammogram.

Some researchers developing a computer-aided system aim at assessing the risk factors, detection and diagnosis of breast cancer using the features found on the mammogram, including: color feature [7], texture [8], [9], shape [10] and a combination among the three [11]. The use of the right features greatly affects the performance of the pattern recognition system. In computation, it is expected to use the features as minimum as possible and to be able to distinguish between one class from another. Therefore, it needs an algorithm that can be used to choose the best features among so many features. Some previous researches have applied several algorithms aimed at the feature selection, among others: the branch and bound algorithm [12], hill climbing algorithm [13] and multi structure co-occurrence descriptor [14]. However, some existing references are not specifically used yet to select the features in the mammogram image for the development of CADx of the breast cancer system.

This research proposes the use of several methods of data mining that are used as the feature selection algorithm of the mammogram image. The algorithms used are the decision tree and the rule induction, afterwards the classification is performed on the features selected from the two algorithms using several classification algorithms to measure the performance. Besides, this research uses the primary data, which types of lesions (benign and malignant) have been classified by the Radiologists not only based on the visual assessment but also verified based on the results of laboratory tests and assessment using other imaging technology that is ultrasound technology.

2. RESEARCH METHOD
This research uses the six-stage process for developing a computer-based system for the diagnosis of breast cancer, including:

2.1. Mammography Image Acquisition
This research uses the primary data in the form of mammogram image produced by digital mammography imaging technology that is conducted in Kotabaru Oncology Clinic Yogyakarta. The number of mammogram image successfully obtained from the probandus is 117 lesions of mammograms form two views, CC (Cranio Caudal) and MLO (mediolateral oblique). Furthermore, the Radiologists in this case as the researchers, conduct a visual analysis of the mammogram. In assessing the mammogram image, the Radiologists do not only interpret the mammogram image, but also match the interpretation result with the interpretation of the image that is the imaging results with other technologies, in this case using ultrasound technology and the results of pathology tests. In the analysis of the mammogram image, the Radiologists need to crosscheck to some test results using other data in order to provide the valid annotations on parts that are considered as the disorders / cancer, hereinafter referred to as RoI (Region of Interest). Besides providing RoI annotation on the mammogram image, the Radiologists classify it into two categories as benign lesions and malignant lesions. Data of 117 mammograms is divided into benign lesions amounted 79 benign mammogram and malignant lesions amounted 38 mammograms. The resulting image of mammography imaging has the same size that is 2424x3296 pixels, but the image of the cropping results, which is the annotations of Radiologists, has the very various sizes because it depends on the level of the vastness of the area of RoI itself.

2.2. Praprocessing
Interpreting the mammogram image is a very difficult job because the image resulting from the mammography technology has a very low quality. One of the characters is having a very low level of contrast that is very difficult to distinguish between the RoI from the fatty tissue. Therefore, before performing the feature extraction, the mammogram image quality needs to be improved, hereinafter referred to as

pretreatment process that aims to get the better qualified image. Some processes performed at this stage include: normalization of mammogram image size to be 256x256 pixels with bilinear interpolation; removing the background of mammogram image with a rolling ball radius of 50 pixels; removing the noise by median filtering with a radius of 2 pixels; improving the image contrast using CLAHE (Contrast-Limited Adaptive Histogram Equalization) method with block size of 127, histogram bins of 256 and maximum slope of 3; besides using CLAHE to improve the image contrast also using equalization histogram with saturated pixel of 0.4%. The results of each stage of pretreatment are shown in Figure 1 and Figure 2.
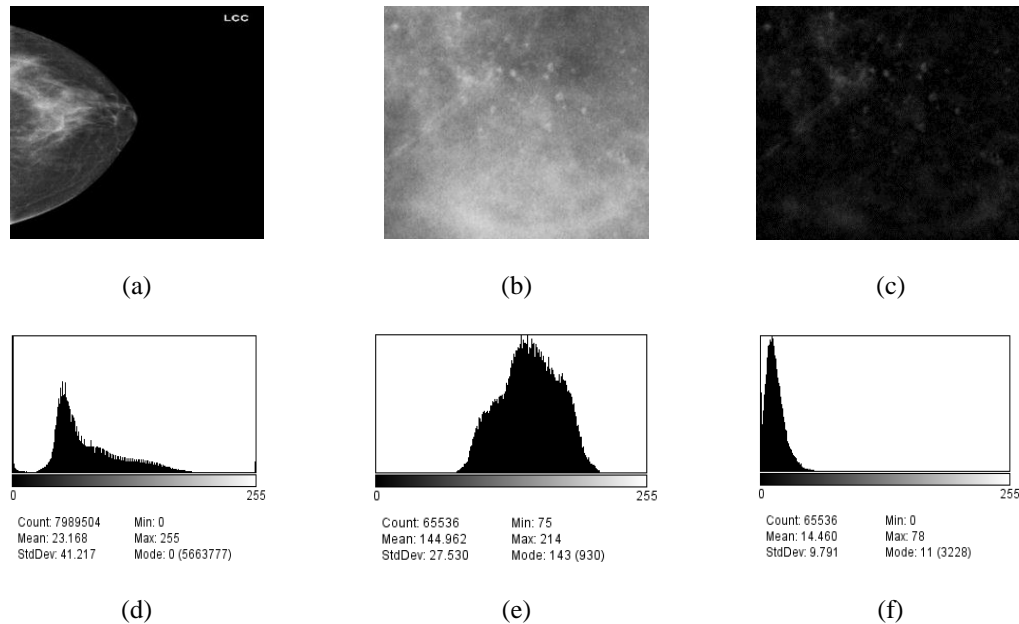

(a)   (b)   (c)

Count: 7989504   Min: 0
Mean: 23.168   Max: 255
StdDev: 41.217   Mode: 0 (5663777)

Count: 65536   Min: 75
Mean: 144.962   Max: 214
StdDev: 27.530   Mode: 143 (930)

Count: 65536   Min: 0
Mean: 14.460   Max: 78
StdDev: 9.791   Mode: 11 (3228)

(d)   (e)   (f)

Figure 1. (a) original image with view CC, (b) Resize RoI, (c) background removal,
(d-f) the histogram

(a)   (b)   (c)

Count: 65536   Min: 0
Mean: 14.303   Max: 72
StdDev: 8.886   Mode: 10 (3819)

Count: 65536   Min: 0
Mean: 41.318   Max: 148
StdDev: 25.183   Mode: 26 (1319)

Count: 65536   Min: 0
Mean: 106.363   Max: 255
StdDev: 62.057   Mode: 65 (404)

(d)   (e)   (f)
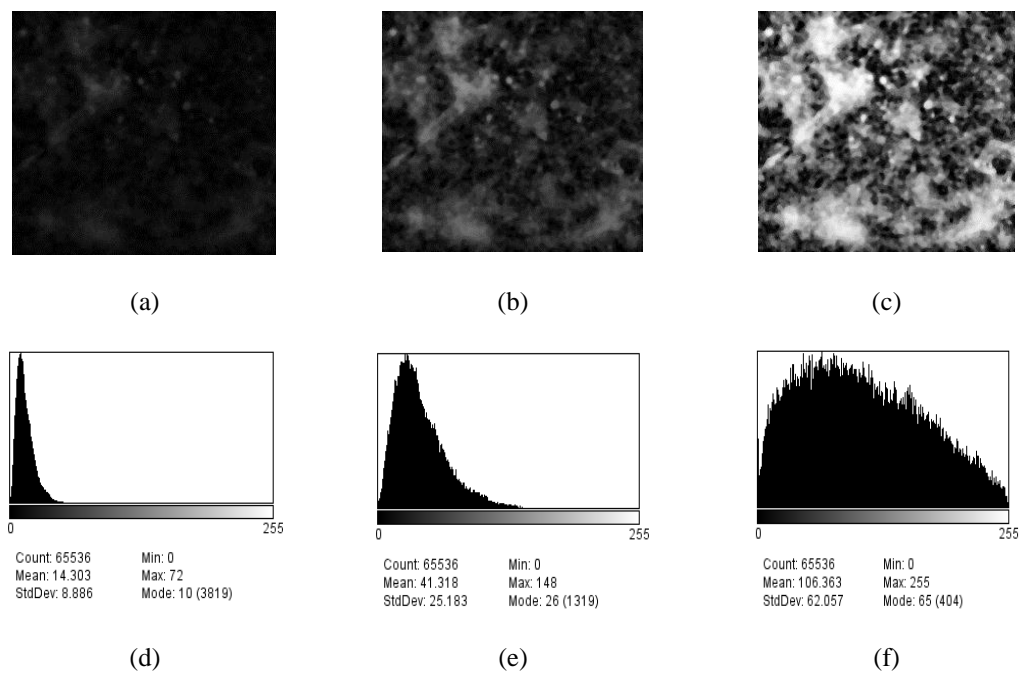
Figure 2. (a) median filtering, (b) CLAHE, (c) equalization histogram, (d-f) the histogram

2.3. Feature Extraction

One key to the success of a pattern recognition system, which in this case is CADx system, is the right use of features that are able to distinguish between benign lesions from malignant lesions. Therefore, it is necessary to study the use of algorithms that can be used to select and evaluate the use of features precisely. In general, this research performs the feature extraction on a mammogram image using two types of feature domain, shape domain (14 descriptors) and texture domain using the equation shown in Table 1 and Table 2. The use of features on the texture domain (24 descriptors) consists of the first order statistics and the second order statistics commonly called GLCM (gray level co-occurrence matrix). The feature of GLCM uses four directios (0^0, 45^0, 90^0 and 135^0) and the average value for each feature with the four directios.

Table 1. Some features for shape domain

Fitur	Expression
Area	Area of selection in square pixels
Center of Massa	the brightness-weighted average of the x and y coordinates all pixels in the image or selection
Modal gray value	the highest peak in the histogram
Centroid	the average of the x and y coordinates of all of the pixels in the image or selection
Perimeter	The length of the outside boundary of the selection
Integrated density	The sum of the values of the pixels in the image or selection
Median	The median value of the pixels in the image or selection
Area fraction	For thresholded images is the percentage of pixels in the image or selection that have been highlighted. For non-thresholded images is the percentage of non-zero pixels
Stack position	The position (slice, channel and frame) in the stack or hyperstack of the selection
Circularity	Circularity = 4π x (area) (perimeter)^2
Aspect Ratio (AR)	The aspect ratio of the particle's fitted ellipse
Roundness	the inverse of *Aspect Ratio*
Solidity	Solidity=$\frac{area}{convex\ area}$

Table 2. Some features for texture domain

Fitur	Expression
Mean (μ)	$\mu = \sum_{k=0}^{255} k \times p(k)$
Standard deviation (σ)	$\sigma^2 = \sum_{k=0}^{255} (k-\mu)^2 \times p(k)$
Skewness	$\sigma = \sum_n (f_n - \mu)\rho(f_n)$
Kurtosis	$\alpha_4 = \frac{1}{\alpha_4} \sum_n (f_n - \mu)^4 p(fn - 3)$
Contrast (CONT)	$CONT\_d\_\theta = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2 \rho_{d,\theta}(i,j)$
Correlation (CORR)	$CORR\_d\_\theta = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{(ixjx\rho_{d,\theta}(i,j) - \mu_x \times \mu_x}{\sigma_x \sigma_x}$
Energy	$Energy\_d\_\theta = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \rho_{d,\theta(i,j)^2}$
Homogeneity (HOMO)	$Homo\_d\_\theta = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{1}{1+(i-j)^2}\rho_{d,\theta}(i,j)$

After praprocessing, the shape and texture domains are extracted on the mammogram image. There are three scenarios of experiments conducted as follows: *first*, using all the features of shape and texture amounted 38 descriptors simultaneously, *second*, using the shape feature with 14 descriptors, and *third*, using the texture feature with 24 descriptors. Scheme of feature use for both domains are shown in Table 3.

Table 3. The use of feature descriptor for experiment

Node	Feature	Deskriptor
1	Shape and texture features	38
2	Shape feature	14
3	Texture feature	24

2.4. Feature Selection

The proper use of the features may provide the optimal classification results, besides, in computation, it may also reduce the burden of processor for unimportant data processing. Therefore, in this research the researchers conduct the data mining as the results of feature extraction with three nodes as noted in Table 3. To perform the feature selection, the mammogram image uses two algorithms those are decision tree and rule induction. Decision tree is a powerful and popular algorithm for classification and prediction. Its other advantage is being able to represent some rules that are easily understood by the humans and the knowledge can be used as the data in the database [16]. While the rule induction algorithm is one of the algorithms implemented on machine learning that is able to formulate some rules extracted from a collection of observation data. The results of data extraction in the form of a rule are the data model in the scientific form that represents some data patterns [17]. The example of the use of decision tree and rule induction for the first node with 38 descriptors is shown in Figure 3 and Table 4.

Some important features are obtained based on the results of mining using decision tree and rule induction for the 38 descriptors of mammogram images. The important features generated by the decision tree algorithm (see Table 5, scenario I) include: kurtosis, area fraction and mean, while the important features generated by the rule induction algorithm (see Table 5, scenario II) include: slice, mean, area fraction and contrast with the angle 135. The same thing is applied to node 2 and 3 using decision tree and rule induction algorithms, in which the mining results are shown in Table 5 (scenario III and IV) for node 2 and Table 5 (scenario V and VI) for node 3. The features used in the experiment with scenarios VII is an important feature generated by the first node using both the decision tree algorithm and the rule induction algorithm. The same thing is applied to scenario VIII and IX using the best features of the experimental results of node II and III. The detailed results of experiment can be seen in Table 5.
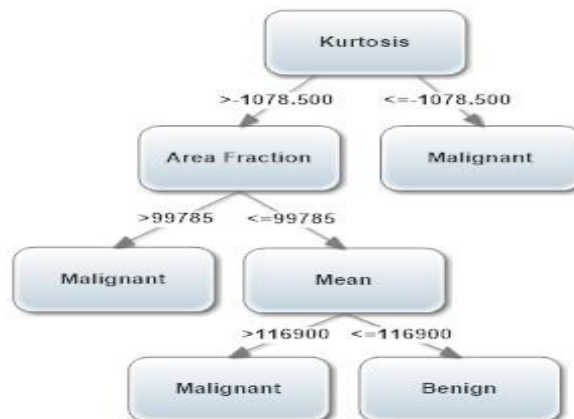

Figure 3. The graph of the result of decision tree algorithm

Table 4. The rule generated by the rule induction for classification of benign dan malignant lesions using 38 descriptors

Node	Rule
1	if Slice > 38500 then Benign (0/41)
2	if Mean > 109600 and AreaFraction ≤ 99730 then Malignant (19/0)
3	if AreaFraction ≤ 99785 and Contrast_135 ≤ 0.373 then Benign (3/30) else Ganas (14/7)

Table 5. The selected features resulting from some scenarios of experiments using decision tree and rule induction algorithms

Scenario	Descriptor	Decision Tree	Scenario	Descriptor	Rule Induction
I	3	Kurtosis, area fraction, mean	II	3	Slice, mean, area fraction, contrast_135
III	6	Area fraction, median, integrated density, slice, modal gray value, center of massa	IV	5	Slice, integrated density, Area fraction, modal gray value, center of massa
V	2	Kurtosis, mean	VI	4	Mean, kurtosis, contrast_0, energy_45, correlation_45
VII	2	Area fraction, mean	VIII	5	Slice, integrated density, Area fraction, modal gray value, center of massa
IX	2	Kurtosis, mean	X	8	Slice, mean, area fraction, contrast_135, integrated density, modal gray value, center of massa, kurtosis

2.5. Classification

Having obtained some of the selected features for each scenario based on the decision tree and rule induction algorithms, then the researchers conduct a classification process of mammogram image into two classes, benign lesions and malignant lesions. In this classification stage, the researchers use several algorithms, among others: k-nearest neighbors (KNN), decision tree (DT) and Naive Bayesian (NB) that further will be expressed in the points of discussion. Based on the feature selection process in the previous process, there will be a classification selection process on ten scenarios predefined previously to measure the performance.

2.6. Evaluation

To evaluate the results of classification of some features based on the selected feature in each scenario, the data is automatically divided using the k-fold cross validation (with 10 k number) in stratified sampling way. Besides, this research also uses five statistical parameters that are commonly used in medical diagnostic result test including: accuracy, sensitivity, specificity, false positive rate (FPR) and true positive rate (TPR). The aim of using the five parameters is to know how reliable and consistent a system to make diagnosis of breast cancer. Accuracy is the amount of data that is successfully predicted correctly by the classification system either negatively or positively, in which the sensitivity is a measure of success of the classification system in identifying the positive data correctly and the specificity is a measure of success of the classification system in identifying the negative data correctly. FPR shows the average of positive cases identified as the wrong one and TPR for the opposite case. Associations between FPR and TPR parameters can be represented graphically that is called the ROC curve. The use of the ROC curves is to assist in making decision in the search for the best model for the diagnosis of breast cancer. The calculation of the five parameters is shown in Table 6.

Illustration: TP (True Positive); TN (True Negative); FN (False Negative); FP (False Positive); false positive rate (FPR) dan true positive rate (TPR).

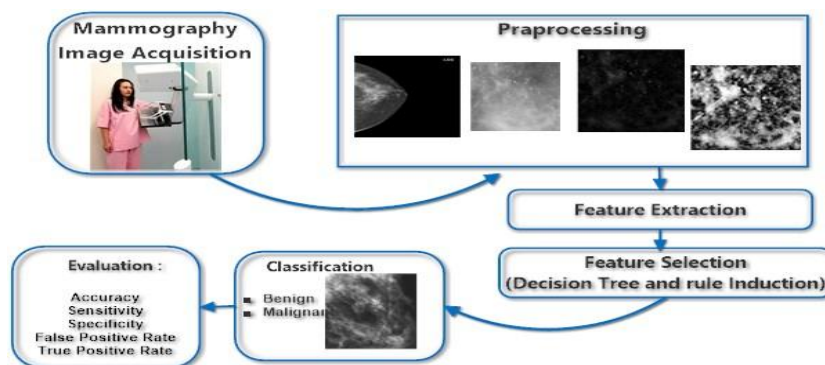General description for each stage is shown in Figure 4.

Figure 4. General description of research stages

Table 6. Formula to obtain the values of sensitivity, specificity dan accurcy

No	Formula
1	$Sensitivity = TP/(TP + FN)$
2	$Specificity = TN/(TN + FP)$
3	$Accuracy = (TN + TP)/(TN + TP + FN + FP)$
4	FPR = FP / (FP+TN)
5	TPR = TP / (TP + FN)

3. RESULTS AND ANALYSIS

The purpose of this research is to find the best features that are used to develop the CADx system for breast cancer on a mammogram image. Therefore, in this research, the researchers have conducted several experiments with ten scenarios, in which each scenario consists of six stages of research that has been described as shown in Figure 4. As an example for experiment with the first scenario, after shooting using the mammography technology, the researchers conducted several times a pretreatment process that has been described in detail in section 2.b. The output of these stages is the obtainment of mammogram images with better quality, so that visually the Radiologists can differentiate between fatty tissue and fat, which previously it was very difficult to distinguish between these two areas because it is a very thin network with no much different intensity. The next stage is to perform the feature extraction of 38 descriptors (a combination of shape and texture features); then the results of the feature extraction are selected using a decision tree (scenario I). The results of the mining process using a decision tree is a fact that not all features are able to contribute in determining the class of breast cancer (benign and malignant). There are only three descriptors that contribute as shown in Table 5. The next process is the stage of mammogram lesion classification into two classes (benign and malignant) using the algorithm of K-Nearest Neighbor (KNN), decision Tree (DT) and Naive Bayesian (NB). A classification is performed in the use of features for each scenario using three classification algorithms and there is an evaluation process using 10-fold cross validation. The complete results for each stage of the evaluation are shown in Table 7. The highest accuracy value is obtained at the CADx system to classify between benign and malignant lesions in scenario IV and VIII (using the five descriptors as shown in Table 5) with the classification algorithm of Decision Tree amounted 93.18%. The use of the five descriptors also provide values of FPR, TPR, Precision and Recall of 6%; 92%; 88% and 92%, while the rule can be used to classify both types of breast cancer as shown in Table 8.

Table 7. Evaluation result of the use of selected features (based on the algorithms of decision tree dan rule induction) in accordance with the scenario

Classification Method	Acc (%)	Sens (%)	Spec (%)	FPR (%)	TPR(%)
SCENARIO I					
KNN	76,29	63,16	17,72	17,72	63,16
DT	82,05	72,97	13,75	12,66	71,05
NB	60,83	44,74	9,76	53,16	89,47
SCENARIO II					
KNN	76,29	62,5	16,88	18,99	65,79
DT	88,26	76,09	4,23	13,92	92,11
NB	77,05	59,02	3,57	31,65	94,74
SCENARIO III					
KNN	71,36	56,25	23,53	17,72	47,37
DT	90,61	81,39	4,05	10,13	92,11
NB	77,12	59,65	6,67	29,11	89,47
SCENARIO IV and VIII					
KNN	71,36	56,25	23,53	17,72	47,37
DT	93,18	87,5	3,89	6,33	92,11
NB	75,3	57,63	6,89	31,65	89,47

Classification Method	Acc (%)	Sens (%)	Spec (%)	FPR (%)	TPR(%)
SCENARIO V and IX					
KNN	77,27	64,87	17,5	16,46	63,16
DT	76,06	72,73	23,16	7,59	42,11
NB	70,15	52,83	15,63	31,65	73,68
SCENARIO VI					
KNN	77,27	64,86	17,5	16,46	63,16
DT	75,23	68	22,83	10,13	44,74
NB	62,5	45,71	12,77	48,10	84,21
SCENARIO VII					
KNN	79,77	67,5	14,29	16,46	71,05
DT	83,03	72,5	11,69	13,92	76,32
NB	61,74	45,57	5,26	54,43	85,71
SKENARIO X					
KNN	71,36	2,35	3,2	17,72	47,37
DT	84,55	77.78	0,52	10,13	73,68
NB	76,21	58.62	0,3	30,38	89,47

Table 8. The rule resulting in the rule induction for the classification of benign and malignant lesions using five descriptors (selected features of the result of scenario IV and VIII)

Node	Rule
1	if Slice > 38500 then Benign (0 / 41)
2	if IntegratedDensity > 7171500000 and AreaFraction ≤ 99730 then Malignant (19 / 0)
3	if AreaFraction ≤ 99785 and AreaFraction > 99705 then Benign (0 / 18)
4	if AreaFraction ≤ 99695 and AreaFraction > 99610 then Benign (0 / 13)
5	if ModalGrayValue > 17500 then Malignant (18 / 2)
6	if Center of Massa ≤ 131650 then Benign (0 / 5)
7	else Malignant (0 / 0)

Classification algorithm (KNN, DT and NB) is used for the development of CADx system with selected features (five descriptors); further, the testing is performed with 10-fold cross validation and visualized using RoC curve as shown in Figure 5. DT algorithm performs best performance when compared to the other two algorithms.
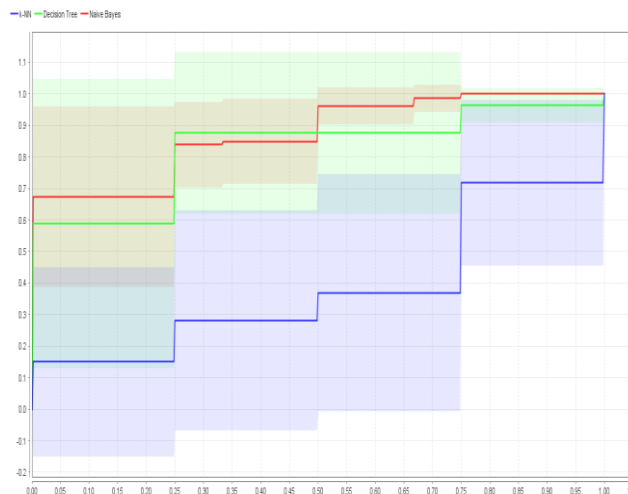

Figure 5. The testing result of feature use: slice, integrated density, Area fraction, modal gray value and center of massa

4.   CONCLUSION

Based on the results of experiments and tests, it can be concluded that decision tree and rule induction, which is one of the mining algorithms, can be used as an alternative method for selection of features on the mammogram image. The results of feature selection conducted with ten scenarios obtain five descriptors that have contributed in the mammogram image classification into two classes, benign lesions and malignant lesions. The best classification results based on the five features (slice, integrated density, Area fraction, gray capital value, center of mass) are generated by the decision tree algorithm with accuracy, sensitivity, specificity, FPR and TPR of 93.18%; 87.5%; 3.89%; 6.33% and 92.11%.

REFERENCES

[1]   Badan Penelitian dan Pengembangan Kesehatan, "Riset Kesehatan Dasar", *Kementerian kesahatan RI*, pp. 85-86, 2013.
[2]   American College of Radiology, "ACR BI-RADS Atlas Fifth Edition", online https://www.acr.org/
[3]   S. Uyun and S. Hartati, "Model Komputasi Penentuan Faktor Resiko Kanker Payudara Berdasarkan Pola dan Persentase Densitas Mamografi", *Doctoral dissertation*, Universitas Gadjah Mada, 2014.
[4]   S. Uyun, *et al.*, "Improvement of Sample Selection: A Cascade-Based Approach for Lesion Automatic Detection," *International Journal of Advanced Computer Science and Applications,* vol. 7, no. 4, pp. 175-182, 2016.
[5]   R. O. Duda, *et al.*, "Pattern classification", John Wiley & Sons, 2012.
[6]   Zheng, *et al.*, "Feature Selection Method Based on Improved Document Frequency," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 12, no. 4, pp. 905-910, 2014.
[7]   M. Langarizadeh and R. Mahmud, "Breast Density Classification Using Histogram-Based Features", *Iranian Journal of Medical Informatics*, vol. 1, no. 1, 2012.
[8]   D. A. Chandy, *et al.*, "Texture feature extraction using gray level statistical matrix for content-based mammogram retrieval", *Multimedia tools and applications*, Vol. *72, No.* 2, pp. 2011-2024, 2014.
[9]   S. Kavitha and K. K. Thyagharajan, "Features based mammogram image classification using weighted feature support vector machine". *In Global Trends in Information Systems and Software Applications,* Springer Berlin Heidelberg, pp. 320-329, 2012.
[10] B. Surendiran and A. Vadivel, "A new feature reduction method for mammogram mass classification". *In Control, Computation and Information Systems*, Springer Berlin Heidelberg, pp. 303-311, 2011.
[11] D. A. Chandy, D. A., *et al.*, "Neighbourhood search feature selection method for content-based mammogram retrieval". *Medical & biological engineering & computing*, vol. 1, no. 13, 2016.
[12] Z. Wang, *et al.*, "An improved branch & bound algorithm in feature selection". *In International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing,* Springer Berlin Heidelberg, pp. 549-556, 2003.
[13] C. M. Nunes, *et al.*, "Feature subset selection using an optimized hill climbing algorithm for handwritten character recognition". *In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, Springer Berlin Heidelberg, pp. 1018-1025, 2004.
[14] Minarno, *et al.*, "Image Retrieval Based on Multi Structure Co-occurrence Descriptor", *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 14, no. 3, pp. 1175-1182. 2016.
[15] G. Alka and R. Patel. "Performance Analysis of SMC Protocols for Decision Tree Classification Rule Mining." *International Journal of Simulation--Systems, Science & Technology,* vol. 13, no.6, 2012.
[16] B. Jerzy, *et al.*, "Sequential covering rule induction algorithm for variable consistency rough set approaches." *Information Sciences,* vol. 181, no. 5, 987-1002, 2011.

BIOGRAPHIES OF AUTHORS

Dr. Shofwatul 'Uyun, S.T., M. Kom is a Full Time Lecturer at the department of Informatics and Head of Information Technology and Database, State Islamic University (UIN) Sunan Kalijaga in Yogyakarta, Indonesia. She obtained her Bachelor degree in Informatics from Islamic University of Indonesia. She received her M.Kom. and Dr in Computer Science from the Gadjah Mada University. Her research interests are pattern recognition, artificial intelligence and medical image processing.

Dr. dr. Lina Choridah, Sp.Rad (K) is Head of Radiology Department, Faculty of Medicine, Universitas Gadjah Mada. She obtained her Bachelor degree (MD/Medical Doctor), Radiologist (Sp.Rad), *Breast Imaging Consultan* (Sp.Rad (K)) and Dr in Medicine from Faculty of Medicine the Gadjah Mada University. Starting in 2004 as the field of Radiology the breast and was a member of team Oncology especially Breast Cancer Faculty of Medicines The Gadjah Mada University.