Journal of Advanced Health Informatics Research (JAHIR)
Vol. 1, No. 1, April 2023, pp. 28-34
DOI: https://doi.org/10.59247/jahir.v1i1.20

Identification of Breast Tumors With Image Processing Using Canny Edge Detection

Deny Nugroho Triwibowo [1], Bala Putra Dewa [2], Bagus Bambang Sumantri [3], Riska Suryani [4]
[1]*Department of Information technology, Universitas Harapan Bangsa, Purwokerto, 53182 Indonesia*
[2] *Department of Informatics, Universitas Harapan Bangsa, Purwokerto, 53182, Indonesia*
[3,4] *Department of Information System, Universitas Harapan Bangsa, Purwokerto, 53182, Indonesia*

ARTICLE INFO

Article history:

Received December 18, 2022
Revised January 03, 2023
Published January 18, 2023

Keywords:

Image processing;
Breast tumor;
Edge Detection;
Canny's method;
SVM;

ABSTRACT

Breast tumor is one of the leading causes of death in women worldwide. The term tumor is often used for all lumps found in the human body. The increase in the number of cases of breast tumors that occur each year is due to the absence of prevention or early detection. The research was carried out by utilizing the development of information technology to create a breast tumor detection system with digital image processing. The application system will process mammogram images to detect edges with the canny method and will be classified using the SVM method. The data used is 176 data obtained from the Kaggle dataset. The test results revealed that 64 patients were classified as having malignant breast tumors (M), and 113 did not have breast tumors (B), with a classification accuracy rate of 95%. From the results obtained, the application system is very good for the early identification of breast tumors.

Corresponding Author:

Deny Nugroho Triwibowo, Universitas Harapan Bangsa, Purwokerto, 53182 Indonesia
Email: denynugroho@uhb.ac.id

1. INTRODUCTION

Breast tumors result from abnormal cell growth, whose growth lies in human breast tissue [1]. The term tumor or neoplasm, in general, is often used for all lumps that can be interpreted as swelling in the human body. Tumor growth can be categorized as malignant or benign [2]. Breast tumor is one of the leading causes of death in women worldwide and ranks first in terms of the most diagnosed cancers in Indonesia [3][4].

According to data from the Global Burden of Cancer Study (Globocan), in 2020, the number of new cases of breast tumors reached 68,858 cases or 16.6% of the total 396,914 new cases in Indonesia [5]. Of these, more than 22,000 people died as a result of breast tumors, an increase from the previous number of cases in 2018 of 58,256 out of a total of 348,809 cases of breast tumors [6][7]. Risk factors that are closely related to the incidence of breast tumors include female sex, while the causes for a woman to develop breast tumors include family and genetic history, previous history of breast disease, history of early menstruation (<12 years), obesity, alcohol consumption, history of chest wall radiation, environmental and behavioural factors [8]. The increase in the number of cases of breast tumors that occur each year is due to the absence of prevention or early detection [9].

One way to detect a lump in the breast early is with the BSE technique or breast self-examination. Breast self-examination (BSE) is performed to detect or identify early the possibility of breast cancer because it is considered cheap in a simple way [10]. However, the SADARI technique is still not optimal in detecting a woman with breast tumors, according to a study conducted by Qonita, which explained that 70% of patients who underwent examination were in the advanced stage category. Therefore, an application system with a

digital image processing method is needed that can be done alone with a smartphone to detect breast tumours early [11].

Research conducted by Riska Nanda to identify FAM (Fibora Adenoma Mammae), a tumor in the category most often found in the breast during the examination. This study utilizes the Sobel method's segmentation process on medical images resulting from ultrasound examinations. The results obtained from using the Sobel method by segmenting medical images are very good for determining the outlines of FAM objects because the outline boundaries can be seen very clearly. So that medical images that are entered into the system are identified properly and accurately, but medical images that have a low level of resolution will be very difficult to determine the outline of the object entered [12].

Research conducted by Atika Maulida with the detection and segmentation of objects identified as breast tumor lesions on mammogram images using the Otsu Thresholding method and feature extraction using the GLCM method. In this study, the data used were six mammogram medical images. These data have gone through the pre-processing stage by removing noise contained in the image and improving image quality. The results of the tests show that image segmentation in the form of GLCM feature values in each image has different values. This is caused by the use of different angle directions, starting from the angle of 0^0, 45^0, 90^0, and 135^0 [13].

Furthermore, research by implementing the canny method and power law transformation to determine the outline of the cervix in detecting early symptoms of cervical cancer by colposcopy. The data used in the test were ten images with *png image format, measuring 200 x 200 pixels. These were pre-processed using power law transformation, namely changing colposcopy images (colour images) to grayscale images (grayscale). The results of the study obtained a percentage of system accuracy of 80%, where the canny method can carry out edge detection processes quickly, which obtains a running time value of 0.0589809 ms and produces a greater level of image similarity, which obtains a Mean Square Error (MSE) value of 11296, 3 [14].

Based on the description of the problem and the methods that have been described, a study was carried out by creating an application system to detect and classify breast tumor levels for women/women. Image detection is carried out using the canny method to detect the edges of the image as the first step in pattern recognition. Furthermore, when the pattern is obtained from the input image, a classification will be carried out using the SVM (Support Vector Machine) method to classify the image as whether it is in the category of a malignant (malignant) or benign (benign) tumor. It is hoped that the existence of a breast tumor detection and classification application system can be a source of information to find out the symptoms of breast tumors as a form of prevention and early detection to make women aware of the importance of breast health [15].

2. METHODS

According to Dawson, there are four research methods most often used by previous researchers: action research, experimental research, case study research, and surveys. This study uses experimental research utilizing a quantitative approach. Experimental research is a type of research that tries to research a relationship using a test [16].

The medical image data used for testing in this study were taken from the Kaggle dataset and the results from previous studies. The dataset is the result of mammogram images of breast tumor patients consisting of 6 mammogram image datasets of breast tumors which will be processed for image processing to detect the edges of breast tumors. Before processing image data, a literature study of several previous studies and a discussion of the symptoms of patients with breast tumors are first carried out [17].

In the research conducted, it will be divided into two main stages, namely processing as the initial and most important step to perform edge detection of the input image data using the Canny method, then classifying based on the results of the edge detection using the SVM classification method [18].]. At the data interpretation stage, mammogram images' detection and classification results were analyzed using the Python programming language. Through this analysis, it will be known whether the detection and classification results can detect breast tumors contained in mammogram images or not and how the results obtained from these values fall into the category of malignant or benign tumors.

3. DISCUSSION

3.1. Edge Detection Using the Canny Method

The canny method is a method for optimal edge detection. The canny method in its processing uses a Gaussian derivative kernel to filter out the initial image's noise to get smooth edge detection results. The first and most important step in this research is the edge detection of mammogram images that are entered into the system using the non-maximum suppression technique. Non-maximum suppression is a process to remove

values in image data that have no maximum value. Furthermore, hysteresis thresholding is carried out using two thresholds T1 (lower threshold) and T2 (upper threshold). If the detection wave is below T1, that point will be set to zero (non-edge). If the detection wave is above T2, it is included in the edge. If the wave is between T1 and T2, it will be set to zero unless there is a path from that point to a point with waves above T2.

Table 1. Canny Detection Object

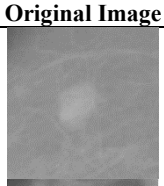
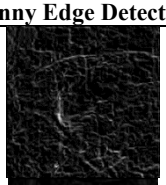
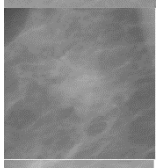
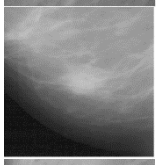
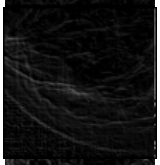
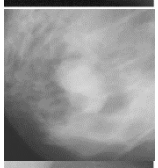
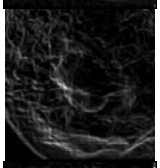
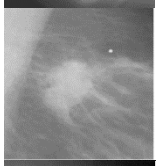
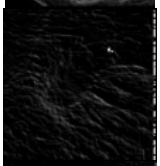
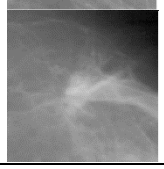
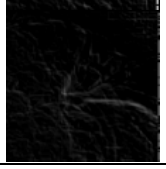
Category	Original Image	Canny Edge Detection	Evaluation Error
Benign Category.1			MSE Canny :112.23668077256946 PSNR Canny :27.629455462498917
Benign Category.2			MSE Canny :69.86184414039262 PSNR Canny :29.68840315360362
Benign Category.3			MSE Canny :110.81363959270661 PSNR Canny :27.684871416733582
Benign Category.1			MSE Canny :110.26306497175142 PSNR Canny :27.706503005151696
Benign Category.2			MSE Canny :79.8521650682941 PSNR Canny :29.10793665008739
Benign Category.3			MSE Canny :93.50368006440114 PSNR Canny :28.42251656945009

Table 1 shows six breast tumor image datasets with category three benign breast tumor image data and three malignant breast tumor image data. From the three benign breast tumor image data, the result is that the "Benign.1" image has the highest MSE (Mean Square Error) value of the three benign breast tumor image data, which means that the degree of similarity between the edge detection data and the original image is very good. Furthermore, from the three malignant breast tumor image data, it is known that the image labelled " Benign.1" has the highest MSE value of the three breast tumor image data with the malignant category, which means that the degree of similarity between the original image and the edge detection results is very good.

3.2. SVM Data Classification

At this stage, breast tumor image data classification has been processed to obtain edge detection using the Canny method. From these data, a breast tumor classification model was created with two types of classification with different ID numbers, namely: number 1 for Malignant and number 0 for Benign. The data used for the classification process was obtained from the Kaggle dataset of 569 data that already has edge feature extraction values. Of the 569 datasets, there are objects or medical images with malignant breast tumors (M) and benign breast tumors (B). The next stage involves more attributes and values in the processed image data to provide useful information when the process is working.

```
radius_mean                0.942380
texture_mean               0.650450
perimeter_mean             0.990650
area_mean                  1.645732
smoothness_mean            0.456324
compactness_mean           1.190123
concavity_mean             1.401180
concave points_mean        1.171180
symmetry_mean              0.725609
fractal_dimension_mean     1.304489
radius_se                  3.088612
texture_se                 1.646444
perimeter_se               3.443615
area_se                    5.447186
smoothness_se              2.314450
compactness_se             1.902221
concavity_se               5.110463
concave points_se          1.444678
symmetry_se                2.195133
fractal_dimension_se       3.923969
radius_worst               1.103115
texture_worst              0.498321
perimeter_worst            1.128164
area_worst                 1.859373
smoothness_worst           0.415426
compactness_worst          1.473555
concavity_worst            1.150237
concave points_worst       0.492616
symmetry_worst             1.433928
fractal_dimension_worst    1.662579
dtype: float64
```

Fig. 1. Data Spread Value

From Fig. 1, radius_mean, perimeter_mean, area_mean, concavity_mean, and concave_points_mean can help predict breast tumour category because of their grouping, which can be distinguished between the types of malignant breast tumors and benign breast tumors. Apart from that, it can also be determined that area_worst and perimeter_worst are useful for predictions. The results of the data distribution show the spread of positive (right) or negative (left) data. Values that are close to zero indicate a smaller spread of slope data. From these results, of the 569 datasets used, the diagnostic results obtained were 212 data in the category of malignant breast tumors (M) and 357 data in the category of benign breast tumors, as shown in Table 2 and Fig. 2.

Table 2. Diagnosis of Breast Tumors

Diagnosis	Amount
Malignant (M)	212
Benign (B)	357

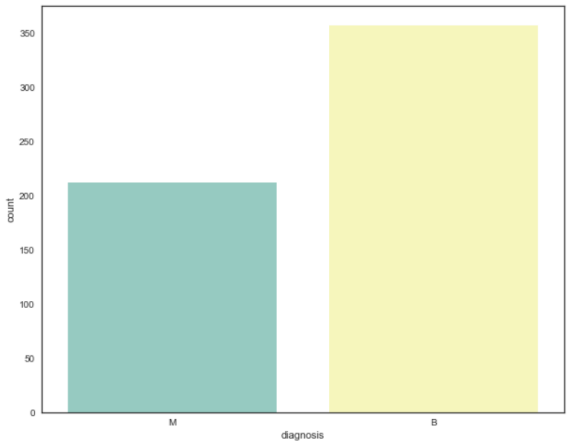


Fig. 2. Breast Tumor Diagnosis Chart

3.3. Model Accuracy

After getting a classification model that has been done, such as in statistical research and machine learning, now to measure the accuracy of the model used for the classification method is the Area Under the Curve (AUC). At this stage, to understand what information is generated by the ROC curve, the calculation is called a confusion matrix, a two-dimensional table with the classification model used on one axis (vertical). Ground truth is on the other axis. (horizontal), as shown in Table 3. Either of the axes can produce two different values.

Table 3. Confusion Matrix

Model says "+"	**Model says "-"**	
True Positive	False Negative	**Actual: "+"**
False Positive	True Negative	Actual: "-"

Re-testing was carried out by entering 174 breast tumor image datasets for edge detection and classifying them according to the category of the data object.

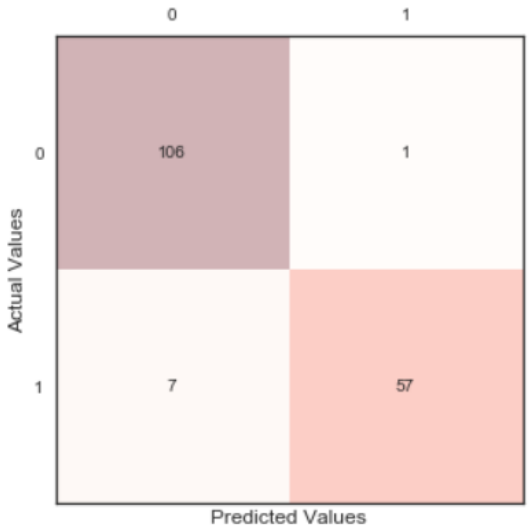


Fig. 3. Classification Testing

Table 4. Classification Results

Category	Precision	Recall	F1-score	Support
0	0.94	0.99	0.96	107
1	0.98	0.89	0.93	64
avg / total	0.95	0.95	0.95	171

From the confusion matrix test results, as contained in Fig. 4 and Table 4, it is known that there are two possible predicted categories, between 1 and 0. Where the number 1 indicates Malignant (indicates the presence of breast tumor cells) with a precision level of 98%, and the number 0 means Benign (indicates the absence of breast tumor cells)) with a precision level of 94%. From the results of the 174 classifications processed, some classifiers predicted "Yes" in as many as 58 data and "No" in 113 data. However, in reality, 64 patients whose medical image data were processed were classified as having breast tumors, and 107 other data did not have breast tumors. The assessment of the confusion matrix results from the classification test shows that the tests carried out have very good accuracy of 95%, followed by other variables, as shown in Table 5.

Table 5. Perhitungan Confusion Matrix

Variable	Calculation	Value
Accuracy	(57 + 106) / 171	0.95
Misclassification Rate	(1+7)/171	0.05
True Positive Rate	57/64	0.89
False Positive Rate	1/107	0.01
Specificity Rate	106/107	0.99
Precision	57/58	0.98
Prevalence	64/171	0.34

4. CONCLUSION

Prevention and detection of breast tumors can be detected earlier in many ways, one of which is by using a digital image processing application system. Digital image processing is done by entering mammogram medical image data for edge detection using the Canny method. The results of edge detection with the Canny method get a fairly high MSE value reaching 112, which means that the similarity between the original image and the edge detection image is very good. After obtaining the results of the edge feature extraction from the Canny method, a classification process is then carried out using the SVM method to categorize the input image data. From the 176 test data, it was found that 64 patients were classified as having breast tumors, and 113 were classified as not having breast tumors, with a classification accuracy rate of 95%. The application system built using the Canny method and the SVM method is very good for the early detection of a person with benign breast tumors or malignant breast tumors.

REFERENCES

[1] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. . Sharma, "Various Types and Management of Breast Cancer : An Review," *J. Adv. Pharm. Technol. Res.*, vol. 1, no. 2, pp. 109–126, 2010, doi: 10.1016/S1474-4422(10)70128-9.

[2] T. Riley, "Benign Breast Disease," *Physician Assist. Clin.*, vol. 3, no. 3, pp. 363–371, 2018, doi: 10.1016/j.cpha.2018.02.005.

[3] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics," *CA. Cancer J. Clin.*, vol. 65, no. 2, pp. 87–108, 2015, doi: 10.3322/caac.21262.

[4] W. H. Organization, "Cancer Today: Data and Methods," *Global Cancer Observatory (GCO)*, 2018. http://gco.iarc.fr/ (accessed Jan. 10, 2023).

[5] C. M. Annur, "Kanker Payudara, Penyakit Kanker Paling Banyak Dialami Masyarakat Indonesia," *Katadata Media Network*, 2022. https://databoks.katadata.co.id/datapublish/2022/10/11/kanker-payudara-penyakit-kanker-paling-banyak-dialami-masyarakat-indonesia (accessed Jan. 10, 2023).

[6] H. Widowati, "Kasus Kanker Payudara Paling Banyak Terjadi di Indonesia," *Katadata Media Network*, 2019. https://databoks.katadata.co.id/datapublish/2019/06/03/kasus-kanker-payudara-paling-banyak-terjadi-di-indonesia (accessed Jan. 10, 2023).

[7] K. K. R. Indonesia, "Kanker Payudara Paling Banyak di Indonesia, Kemenkes Targetkan Pemerataan Layanan Kesehatan," *Kementrian Kesehatan Republik Indonesia*, 2022. https://www.kemkes.go.id/article/view/22020400002/kanker-payudara-paling-banyak-di-indonesia-kemenkes-targetkan-pemerataan-layanan-kesehatan.html (accessed Jan. 13, 2023).

[8] R. F. Ahsani and P. B. Machmud, "Hubungan Riwayat Reproduksi dengan Tumor Payudara pada Perempuan Usia

Muda di Indonesia (Analisis Riset PTM 2016)," *Media Kesehat. Masy. Indones.*, vol. 15, no. 3, p. 237, 2019, doi: 10.30597/mkmi.v15i3.6278.

[9] E. Irawan, L. Rahayuwati, and D. I. Yani, "Hubungan Penggunaan Terapi Modern dan Komplementer terhadap Kualitas Hidup Pasien Kanker Payudara," *J. Keperawatan Padjadjaran*, vol. 5, no. 1, pp. 19–28, 2017, doi: 10.24198/jkp.v5n1.3.

[10] A. C. Sari, "Pengaruh Pendidikan Kesehatan Tentang Pemeriksaan Payudara Sendiri Terhadap Pengetahuan Dan Motivasi Melakukannya Pada Wanita Usia 30-50 Tahun Di Desa Joho Mojolaban," Universitas Muhammadiyah Surakarta, 2016.

[11] Q. Fauziah and A. R. Mulyana, "Perancangan Aplikasi Kesehatan Payudara Sebagai Media Informasi dan Edukasi," *FAD*, pp. 1–14, 2022.

[12] R. Nanda, S. W. Dari, and A. Ihsan, "Segmentasi Citra Medis untuk Deteksi Objek FAM pada Payudara Menggunakan Metode Sobel," *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 248, 2019, doi: 10.30865/mib.v3i4.1232.

[13] A. Maulida, N. Nurhidayah, Y. Fendriani, and H. Haryono, "Segmentasi Citra Mammogram Untuk Deteksi Dini Kanker Payudara Dengan Menggunakan Metode Otsu Thresholding," *J. Fis. Unand*, vol. 11, no. 2, pp. 180–186, 2022, doi: 10.25077/jfu.11.2.180-186.2022.

[14] N. B. Tsani and H. Harliana, "Implementasi Deteksi Tepi Canny Dengan Transformasi Powerlaw Dalam Mendeteksi Stadium Kanker Serviks," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 1, no. 01, pp. 22–33, 2019, doi: 10.46772/intech.v1i01.35.

[15] U. L. Lubis, "Pengetahuan Remaja Putri Tentang Pemeriksaan Payudara Sendiri (SADARI) dengan Perilaku Sadari," *J. Aisyah J. Ilmu Kesehat.*, vol. 2, no. 1, pp. 81–86, 2017, doi: 10.30604/jika.v2i1.36.

[16] C. W. Dawson, *Projects in Computing and Information Systems*, vol. 2. 2009. [Online]. Available: http://www.sentimentaltoday.net/National_Academy_Press/0321263553.Addison.Wesley.Publishing.Company.Projects.in.Computing.and.Information.Systems.A.Students.Guide.Jun.2005.pdf

[17] N. U. Maharani, "Gambaran Penderita Tumor Payudara Berdasarkan Usia Biologis," *J. Med. Hutama*, vol. 03, no. 02, pp. 1851–1854, 2022.

[18] M. I. Rosadi, A. Z. Arifin, and A. Yuniarti, "Klasifikasi Massa Pada Citra Mammogram Menggunakan Kombinasi Seleksi Fitur F-Score Dan Ls-Svm," *Teknol. J. Ilm. Sist. Inf.*, vol. 6, no. 1, p. 27, 2016, doi: 10.26594/teknologi.v6i1.558.