Journal of Holistic Medical Technologies
Vol.1 No.1 (2024) 40 – 49
Available online at https://journal.innoscientia.org/index.php/jhmt

Classification of normal and cancerous mammogram images based on texture features using the Support Vector Machine (SVM) method

Risma Eka Ashari[1*] and Hamdan Hadi Kusuma[1]
[1] Department of Physics, Faculty of Science and Technology, Universitas Islam Negeri Walisongo Semarang, Indonesia
*Corresponding author's e-mail: risma_eka_ashari_2008026001@walisongo.ac.id

ABSTRACT

Breast cancer is the leading cause of death in older women, with more than one million women worldwide dying from this disease yearly. Mammography is a specialized radiological examination that uses low-dose X-rays to detect breast abnormalities, even before visible symptoms such as palpable lumps appear. This study aims to develop an effective mammogram image classification model using the SVM (Support Vector Machine) method with texture feature extraction analysis in histograms and GLCM (Gray-Level Co-Occurrence Matrix). The research involved 20 normal and 20 cancer images, starting with mammogram image preprocessing, then texture feature extraction using histograms and GLCM, and ending with data classification using the SVM method. Test results showed that SVM could classify images with an accuracy of 67.5%, a sensitivity of 33.3%, and a specificity of 70%. Therefore, this research could be a foundation for further developments to enhance mammogram image classification accuracy.

Keywords:
Cancer; SVM; Texture Feature; GLCM; Mammogram

Introduction

Breast cancer is the leading cause of death among older women. According to estimates from the International Cancer Research Institute of the World Health Organization, more than one million women die from breast cancer worldwide each year (Rubio et al., 2015). In Indonesia, breast cancer is one of the most prevalent types of cancer, with 28.7% of cancer patients who have breast cancer, based on the Hospital Information System (SIRS) (Retno Paras Rasmi, 2020). Breast cancer is a disease that many women fear (Ullah et al., 2021). One of the most effective methods for detecting and diagnosing breast cancer is through mammography examination using X-rays. The resulting image is a mammogram (Listia & Harjoko, 2014). Early detection of breast cancer through mammography can increase the chances of survival (Kele et al., 2011).

Mammography is a radiological examination that uses low-dose X-rays to detect abnormalities in the breast, even before visible symptoms such as palpable lumps occur (Hartadi et al., 2011). This examination results in a grayscale image of the breast area, known as a mammogram. Digital mammography analysis is performed as the first step in the early detection of breast cancer. Previously, radiologists analyzed mammograms manually, but with the advancement of technology, digital image processing can be utilized to obtain more accurate results. One of the advantages of feature extraction is that it can increase the efficiency and speed of data processing by eliminating unnecessary or redundant data. Furthermore, feature extraction can be used for accurate classification. It can help identify important patterns and information in data, making it easier to analyze and interpret the data (Qayyum & Basit, 2017).

Morphological feature extraction involves processing images based on the shape of segments or regions within the image. Several morphological characteristics can be used, including area, perimeter, metric, and eccentricity. Another feature extraction method that can distinguish between objects is texture feature extraction, which involves extracting image features based on first-order statistics such as mean, variance, skewness, and kurtosis. First-order texture measurements use statistical calculations based solely on the original image pixel values, such as variance, without considering the relationship between neighboring pixels. On the other hand, the Gray-Level Co-Occurrence Matrix (GLCM) is a method of image feature extraction that utilizes second-order statistical functions such as contrast, correlation, energy, and homogeneity. In second-order texture measurement, the relationship between pairs of original image pixels is considered (Kadir & Susanto, 2013).

In a previous study by Listia & Harjoko (2014), it was shown that the 4-way GLCM feature extraction (0°, 45°, 90°, 135°) with a distance of d=1 had the best accuracy in classifying mammograms, with an accuracy rate of 81.1%. Specifically, the direction of 0° achieved a classification accuracy of 100%. Another study by Syifa et al. (2016) investigated texture analysis of microscopic images of lung cancer based on GLCM features and wavelet transform using the naive Bayes classification method. The feature extraction results using the wavelet transform method showed an accuracy rate of 71.42%, while the GLCM method showed an accuracy rate of 80%. Nugroho (2015) conducted a study on the classification of thyroid nodules based on the textural features of ultrasound images using MATLAB. This study used three textural methods to classify thyroid nodules on ultrasound images, and the classification process was carried out using MLP (Multi-Layer Perceptron), resulting in an accuracy of 86.1%. Novianti & Purnami (2012) conducted a classification process using SVM, resulting in an accuracy of 94.34%, while logistic regression achieved 84.90% accuracy using an 80:80 partition. This shows that SVM is better than logistic regression in terms of classification accuracy. Lussiana et al. (2013) successfully classified mammogram images by taking three GLCM feature values using SVM, achieving an accuracy of 90% for the normal category and 87.67% for the abnormal category. The results of this process produce the same output as the label given to the training data feature, with label 0 indicating the condition of the mammogram image for the normal category and label 1 indicating the suspected mammogram image or abnormal category. Marlina et al. (2020) used the Support Vector Machine (SVM) algorithm to classify mammographic images of the breast and achieved an accuracy of up to 90% with 25 normal images and 25 abnormal images. These studies show that SVM has fairly good accuracy in the classification process, with most studies using GLCM as a feature extraction analysis.

Based on this explanation, the study applies SVM to classify mammogram images using texture feature extraction analyses, namely histograms and GLCM. The classification process uses SVM with the assistance of machine learning Weka to group data into two classes: normal and cancer.

This study aims to create an effective mammogram image classification model utilizing the SVM method with texture feature extraction analyses in the form of histograms and GLCM. The model will be used to differentiate between normal and cancer mammogram images. It is hoped that the findings of this study will aid in enhancing the accuracy of breast cancer diagnosis by leveraging medical image processing technology.

Methods

In this study, there are 20 normal images and 20 cancer images were used for analysis. The mammogram images were obtained from a freely available database at http://www.eng.usf.edu/cvprg/mammography/database.html. Texture feature extraction was performed using Matlab 2017, and classification was carried out using Machine Learning WEKA 3.9.6. The research procedure consisted of three stages, as illustrated in Figure 1: mammogram image preprocessing, texture feature extraction using histograms and GLCM, and the final stage of data classification using the SVM method.

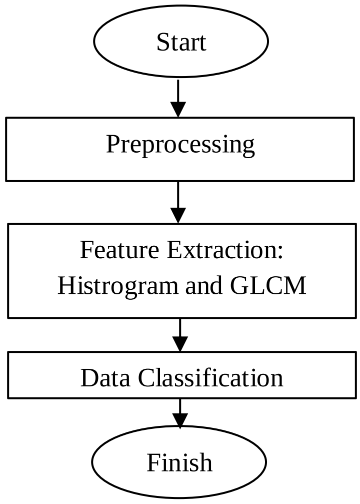


Figure 1. Research Procedure

The preprocessing stage consists of four stages: image retrieval from the DDSM (Digital Database for Screening Mammography), filtering to remove markers and convert RGB images to grayscale, cropping to remove unnecessary parts, and image resizing to improve the image size to 512 x 512 pixels.

Texture feature extraction is conducted using two methods, namely histogram and GLCM. Histogram is a process of histogram equalization, which aims to even out the distribution of gray-level values in an image. In order to perform histogram equalization, a cumulative distribution function is required. This study used six histogram features, including mean, standard deviation, variance, entropy, slope, and kurtosis, as shown in the following equation (Maesyaroh, 2022).

1. Mean is a feature that calculates the average brightness of objects.
$$m = \sum_{I=0}^{L-1} i.p(i) \tag{1}$$
Where: m = Average intensity, i = Gray level in the image $p(i)$ = Probability of occurrence of i and L, iInd L = Highest gray level value.

2. Standard deviation (σ) is a feature that describes the degree of variation or spread of data from the mean value of a measurement.
$$\sigma = \sqrt{\sum_{I=0}^{L-1} (i-m)^2.p(i)} \tag{2}$$

3. Variance is a feature that provides information on the size of the contrast in an image
$$the\ \sigma^2 = \sqrt{\sum_{I=0}^{L-1} (i-m)^2 p(i)} \tag{3}$$

4. Entropy is a feature that describes the degree of randomness or uncertainty in an image
$$Entropi = \sqrt{\sum_{I=0}^{L-1} p(i) log_2(p(i))} \tag{4}$$

5. Skewness is a feature that measures the deviation of a distribution from its average intensity.

$$the\ Skewness = \sqrt{\sum\nolimits_{I=0}^{L-1} (i-m)^3 p(i)} \tag{5}$$

6. Kurtosis is a feature that describes the degree of peakedness or flatness of the histogram curve.

$$Kurtosis = \sqrt{\sum\nolimits_{I=0}^{L-1} (i-m)^4 p(i) - 3} \tag{6}$$

GLCM is used because this method works for grayscale image data. GLCM is a joint probability distribution of grey levels in pixel pairs that satisfy a certain relative position in an image. GLCM uses four features, including Energy, Contrast, Correlation, and Homogeneity, as shown in the following equation (Anggoro, 2016)

1. Energy is a measure of the homogeneity of an image.

$$Energi = \sum\nolimits_i \sum\nolimits_j p^2(i,j) \tag{7}$$

2. Contrast measures the variation between the degrees of gray in the image.

$$Kontras = \sum\nolimits_i \sum\nolimits_j (i-j)^2 p(i,j) \tag{8}$$

3. Correlation is a linear measure of the degree of gray-level interdependence in an image.

$$Korelasi = \frac{1}{\sigma_x \sigma_y} \sum\nolimits_i = 1 \sum\nolimits_j = 1(j - \mu_y) p(i,j) \tag{9}$$

4. Homogeneity is a measure of the uniformity of grey-level transitions in the image.

$$Homogenitas = \sum\nolimits_i \sum\nolimits_j \frac{p(i,j)}{1 + |i-j|} \tag{10}$$

The Support Vector Machine (SVM) method for data classification was first introduced by Vapnik in 1992 during the Annual Workshop on Computational Learning Theory (El Morr et al., 2022).

The basic principle of SVM is that it is a linear classifier, meaning that the t can classify cases that can be separated linearly. For example, given a set $x = \{X_1, X_2, \ldots., X_n\}$, it is declared as a positive class if f(x) ≥ 0, while the others belong to the negative class.

SVM classifies the set of training vectors in the form of paired data sets from two classes (Brereton & Lloyd, 2010).

$$(X_i, y_i), X_i \in R^n, y_i \in \{1, -1\}, i = 1, \ldots., n, \tag{11}$$

The separation of hyperplanes with canonical forms follows the constraints below.

$$y_i[(W^T X_i) + b] \geq 1, i = 2,3, \ldots, n. \tag{12}$$

The optimal hyperplane is obtained by maximizing $\frac{2}{\|W\|}$ or minimizing $\varphi(W) = \frac{1}{2}\|W\|^2$. This optimization problem can be solved using the Lagrange function below.

$$L(w,b,a) = \frac{1}{2}\|W\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i(w^T x_i + b) - 1\right] \tag{13}$$

Here, α_i represents the multiplier of the Lagrange function. Equation (9) represents the primal space, which needs to be transformed into a dual space to make it easier and more efficient to solve. The dual problem can be expressed as follows.

$$\hat{\alpha} = \boldsymbol{arg\ min}\ \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i\alpha_j y_i y_j\ ({x_i}^T x_j) - \sum_{i=1}^{n} \alpha_i \qquad (14)$$

The constraints for the dual problem are α≥0 and $\alpha \geq 0, i = 1$ dan $\sum_{i=1}^{n} \alpha_i\, y_i = 0$
In cases where data cannot be separated, misclassification may occur, so the objective function and constraints are modified by including the slack variable ξ > 0. The formulation becomes as follows.

$$\boldsymbol{\phi(w,\xi)} \frac{1}{2}\|W\|^2 + c\sum_{i=1}^{n} \xi_i \qquad (15)$$

with constraints

$$y_i[(w^T x_i) + b] + \xi_i \geq 1, i = 2, \dots, n. \qquad (16)$$

The difference between the separable and non-separable cases lies only in the addition of the constraint $0 \leq \alpha_i \leq C$ in the non-separable problem.
The optimization of equation (10) in the case of non-linearity becomes as follows,

$$\boldsymbol{and}\ \hat{\alpha} = \boldsymbol{arg\ min}\ \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i\alpha_j y_i y_j\ K(x_i, x_j) - \sum_{i=1}^{n} \alpha_i \qquad (17)$$

with constraints $0 \leq \alpha_i \leq C, i = 1, \dots, n.$ and $\sum_{i=1}^{n} \alpha_i\, y_i = 0$
To handle non-linear data, a kernel function $K(x_i, x_j)$ is used. Based on the steps described in the linear case, the following function is obtained:

$$f(x) = sign\left(\sum_{i=1}^{n} y_i\hat{\boldsymbol{\alpha}}_i(\phi(x_i), \phi(x_j)) + \hat{\boldsymbol{b}}\right) \qquad (18)$$
$$= sign\left(\sum_{i=1}^{n} y_i\hat{\boldsymbol{\alpha}}_i(K(x_i, x_j)) + \hat{\boldsymbol{b}}\right)$$

The sign function labels all values $f(x) < 0$ abel -1 and $f(x) > 0$ as +1. Common kernel functions used in SVM literature include
a. Linear Kernel: $(x^T x)$
a. Polynomial Kernel: $(x^T x + 1)^{\,p}$
b. RBF Kernel: $K(x, y) = exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$
(Armaghani et al., 2020)

The k-Fold Cross-Validation (kFCV) technique is a validation method that involves dividing the dataset D randomly into k subsets, which are independent of each other and denoted as. $f_1, f_2, \dots, f_k$ such that each fold contains one $\frac{1}{k}$ parts of data. Then, k datasets $D_1, D_2, \dots, D_k$ are created, each containing $(\mathrm{k}-2)$ folds for training data, one fold for validation, and one fold for test data (Agussationo et al., 2018).

Furthermore, an analysis was conducted on the test results during the classification stage to evaluate the classification accuracy obtained from the designed test scenarios. The analysis included measuring the number of datasets that were classified correctly, datasets that were misclassified, accuracy values, and benefits (Julia et al., 2022). SVM was originally designed as a linear classifier but was later developed to solve non-linear cases by utilizing the kernel concept in a higher dimensional space. With this kernel concept, SVM can map data into higher dimensions, allowing it to solve cases that cannot be solved by linear classification (A. S. Nugroho, A. B. Witarto, 2003). The success of the classification process can be demonstrated using the measurement index value obtained from the data classification results using WEKA machine learning in the form of a confusion matrix (Ermawati, 2020). The confusion matrix includes True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) values. TP represents images with a normal category classified as normal in the system, while FP represents images with a cancer category classified as normal. TN represents images with a cancer category classified as cancer in the system, and FN represents images with a normal category classified as cancer in the system (Frank, E., Hall, M. A., & Witten, 2017).

Olanyi et al. (2017) proposed using a confusion matrix to measure the performance of a classification method. The confusion matrix includes the accuracy value, which measures the model's ability to classify data correctly. Sensitivity measures the ability of the SVM algorithm to identify mammogram images that are positive for breast cancer correctly. Specificity measures the ability of the SVM algorithm to correctly identify mammogram images that are negative or show no signs of breast cancer. Higher values for accuracy, sensitivity, and specificity indicate better performance of the SVM algorithm in classifying mammogram images, which can be expressed as follows (M. Nuruddin Qaisar Bhuiyan, M. Shamsujjoha, S. H. Ripon, F. H. Proma, 2019):

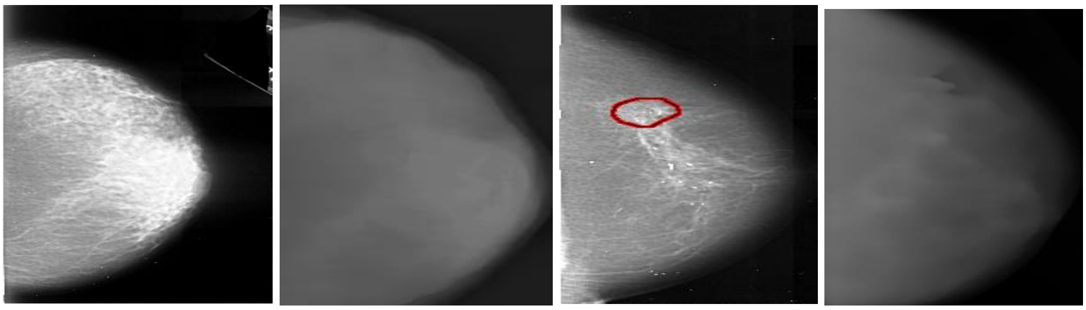
$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{19}$$

$$Sensitivitas = \frac{TP}{TP + FN} \times 100\% \tag{20}$$

$$Spesifitas = \frac{TN}{TP + FN} \times 100\% \tag{21}$$

Results and Discussions

Based on the study's results, the initial stage was preprocessing, which included cropping as the initial process to determine the nodule area. After that, the image underwent a filtering stage to enhance image quality, convert the image to grayscale, and remove markers. Figure 2 shows one of the mammogram images of normal and cancerous breasts before and after preprocessing. The preprocessing was successful, as evidenced by the removal of unnecessary parts, the conversion of the image to grayscale, and the disappearance of the marker.

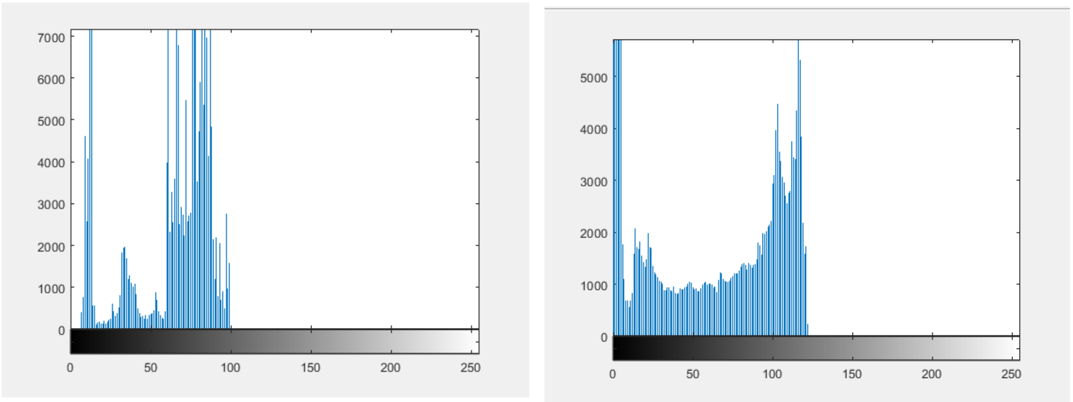


(a) (b) (c) (d)

Figure 2. Mammographic Image a) Normal Before Preprocessing, b) Normal After Preprocessing, c) Malignant Cancer Before Preprocessing, d) Malignant Cancer After Preprocessing

Figure 3 presents the histogram values of normal and cancer mammographic images. Comparing the two histograms reveals different graphical patterns between the two types of images. The graphic pattern appears higher in normal images, resulting in a darker image with values ranging from 60-98.

On the other hand, in malignant images, the graphic pattern shows a lighter image with values ranging from 100-125, decreasing from high to low values.



(a) (b)

Figure 3. shows the histogram of both a) normal and b) malignant cancer images.

Table 1. Average Histogram and GLCM Texture Features

Attribute	Normal	Malignant Cancer
Mean	$31045.7304 \pm 135325.201$	0 ± 2723.3097
Std dev	28.9911 ± 6.0566	34.8567 ± 5.7609
Variants	$4640529.0867 \pm 6482555.1934$	$8928612.9262 \pm 7820702.7654$
Kurtosis	2.2 ± 0.8914	2.1088 ± 0.7371
Skewness	-0.3182 ± 0.4297	-0.6206 ± 0.4249
Entropy	5.3552 ± 0.4371	5.8736 ± 0.3551
Contras	0.0176 ± 0.0047	0.0095 ± 0.0021
Homogeneity	0.9914 ± 0.0024	0.9952 ± 0.001
Energy	0.3501 ± 0.0489	0.3599 ± 0.1025
Correlation	0.9888 ± 0.0052	0.995 ± 0.0015

The texture feature extraction resulted in 10 features, six obtained from the histogram and four from the GLCM. The histogram values included the mean, standard deviation, variance, skewness, kurtosis, and entropy, while the GLCM values included energy, contrast, correlation, and homogeneity.

Table 1 presents the average values of the textural features. In mammographic images, the difference in mean values between malignant and benign tumors indicates variations in textural features. The standard deviation is added to the average value of the texture features to obtain information about the distribution of texture feature data. The standard deviation is an important indicator in determining image clarity. It can be used to evaluate image contrast, where an image with low contrast has a low standard deviation and vice versa (Agussationo et al., 2018). Skewness, contrast, correlation, energy, and homogeneity have uniform data distribution, while the mean, variance, entropy, standard deviation, and kurtosis have random data distribution.

The classification stage is the final stage in image processing. It involves using the Support Vector Machine (SVM) method with the help of Weka's machine learning. The results of the confusion matrix are shown in Table 2. The feature extraction data is provided in CSV format, making it compatible with Weka's machine learning. The input data used for classification is the combined data from texture feature extraction, including histograms and GLCM, with 40 images.

The results of this study show an accuracy value of 67.5%, a sensitivity value of 33.3%, and a specificity value of 70%. These values indicate that the accuracy achieved in this study is lower than that of a previous study conducted by Marlina et al. (Marlina et al., 2020), which used the Support Vector Machine (SVM) algorithm to classify mammographic images of the breast and achieved an accuracy of up to 90% with 25 normal images and 25 abnormal images. The lower accuracy in this study can be attributed to several factors, such as the quality of the mammogram images used, which can affect the SVM's ability to distinguish between normal and abnormal categories. Additionally, the accuracy can be affected by the characteristics of the data, image processing and feature extraction methods, and the specific SVM algorithm used.

Table 2. Confusion Matrix Results

Parameter	Results
TN	7 Data
TP	20 Data
FN	0 Data
FP	12 Data
Accuracy	67,5%
Sensitivity	33,3%
Specificity	70%

The drawback of this study is that it still uses manual methods for preprocessing and feature extraction on mammogram images, so it is likely to be vulnerable to human error and subjectivity tendencies and requires more time. Therefore, it is necessary to develop an automation method.

This research significantly benefits developing breast mammogram image classification technology using the SVM algorithm. Although the resulting accuracy is still lower than previous research, this research can be a basis for further development of factors that can improve the accuracy of mammogram image classification, such as improving image quality, developing better feature extraction methods, or using the SVM algorithm, which is more optimal.

Conclusion

The test results indicate that SVM can classify images with an accuracy rate of 67.5%, a sensitivity rate of 33.3%, and a specificity rate of 70%. Hence, this research can serve as a foundation for further improving the factors that can enhance the accuracy of mammogram image classification, such as enhancing image quality and developing better feature extraction techniques. Through further development, it is hoped that the results of mammogram image classification can become more precise and aid in the early detection of breast cancer.

Acknowledgments

We sincerely thank the Universitas Islam Negeri Walisongo Semarang for its support and facilities, which were crucial for this research. Additionally, we appreciate those who provided the mammogram data essential to this study. May these contributions advance scientific knowledge.

Conflicts of interest

The authors affirm that they have no conflicts of interest.

References

A. S. Nugroho, A. B. Witarto, dan D. H. (2003). Support Vektor Machine. In *Teori dan Aplikasi dalam Bioinformatika*.

Agussationo, Y., Soesanti, I., & Najib, W. (2018). Klasifikasi Citra X-Ray Diagnosis Tuberkulosis Berbasis Fitur Statistis. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *2*(3), 736–745. https://doi.org/10.29207/resti.v2i3.523

Anggoro, W. (2016). *Metadata, citation and similar papers at core.ac.uk Dokumen Karya Ilmiah | Skripsi | Prodi Teknik Informatika - S1 | FIK | UDINUS | 2016*. 0–1.

Armaghani, D. J., Asteris, P. G., Askarian, B., … Huynh, V. Van. (2020). Examining hybrid and single SVM models with different kernels to predict rock brittleness. *Sustainability (Switzerland)*, *12*(6), 1–17. https://doi.org/10.3390/su12062229

Brereton, R. G., & Lloyd, G. R. (2010). Support Vector Machines for classification and regression. *Analyst*, *135*(2), 230–267. https://doi.org/10.1039/b918972f

El Morr, C., Jammal, M., Ali-Hassan, H., & El-Hallak, W. (2022). Support Vector Machine. *International Series in Operations Research and Management Science*, *334*, 385–411. https://doi.org/10.1007/978-3-031-16990-8_13

Ermawati, E. (2020). Klasifikasi Nodul Payudara Berdasarkan Ciri Tekstur pada Citra Ultrasonografi Menggunakan Scilab. *Universitas Islam Negeri Walisongo Semarang*.

ETP, L., Widodo, S., Karmilasari, … Reza, F. (2013). Klasifikasi Citra Mammogram berbasis Tekstur Menggunakan SVM. *Forum Informatika Kesehatan Indonesia (FIKI 2013)*.

Frank, E., Hall, M. A., & Witten, I. H. (2017). *The WEKA workbench. Data Mining*. https://doi.org/https://doi.org/10.1016/b978-0-12-804291-5.00024-6

Hartadi, R., Santoso, I., & Hidayatno, A. (2011). *Deteksi Potensi Kanker Payudara Pada Mammogram Menggunakan Metode Gray Level Co-Occurrence Matrices*. 1–6.

Julia, B., Sumarti, H., & Kusuma, H. H. (2022). *Klasifikasi Tumor Payudara Berbasis Ciri Tekstur pada Citra Mammografi Menggunakan Metode Naive Bayes*. *7*(2), 165–176.

Kadir, A., & Susanto, A. (2013). *Teori dan aplikasi pengolahan citra / Abdul Kadir dan Adhi Susanto*.

Kele, A., Kele, A., & Yavuz, U. (2011). Expert system based on neuro-fuzzy rules for diagnosis breast cancer. *Expert Systems with Applications*, *38*(5), 5719–5726. https://doi.org/10.1016/j.eswa.2010.10.061

Listia, R., & Harjoko, A. (2014). Klasifikasi Massa pada Citra Mammogram Berdasarkan Gray Level Cooccurence Matrix (GLCM). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *8*(1), 59. https://doi.org/10.22146/ijccs.3496

M. Nuruddin Qaisar Bhuiyan, M. Shamsujjoha, S. H. Ripon, F. H. Proma, and F. K. (2019). Transfer learning and supervised classifier based prediction model for breast cancer. *N Big Data Analytics for Intelligent Healthcare Management, Elsevier*, 59–86. https://doi.org/https://doi.org/10.1145/3428757.3429139

Maesyaroh, U. (2022). *Klasifikasi Tumor Payudara pada Citra Mammografi Dengan Metode Support Vector Machine (SVM) Berbasis Ciri Tekstur Statistik*. 121.

Marlina, L., Munadi, K., & Arnia, F. (2020). Klasifikasi Kanker Payudara Menggunakan Teknik Support Vector Machine (Svm) Pada Citra Termografi. *Jurnal Komputer, Informasi Teknologi, Dan Elektro*, *5*(2), 11–13. https://doi.org/10.24815/kitektro.v5i2.15571

Novianti, F. A., & Purnami, S. W. (2012). Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. *Jurnal SAINS Dan Seni ITS*, *1*(1), D147–D152.

Nugroho, A. (2015). *Klasifikasi Nodul Tiroid Berbasis Ciri Tekstur pada Citra Ultrasonografi*.

Olaniyi, E. O., Adekunle, A. A., Odekuoye, T., & Khashman, A. (2017). Automatic system for grading banana using GLCM texture feature extraction and neural network arbitrations. *Journal of Food Process Engineering*, *40*(6). https://doi.org/https://doi.org/10.1111/jfpe.12575 Purwaningsih,

Qayyum, A., & Basit, A. (2017). Automatic breast segmentation and cancer detection via SVM in mammograms. *ICET 2016 - 2016 International Conference on Emerging Technologies*. https://doi.org/10.1109/ICET.2016.7813261

Retno Paras Rasmi. (2020). *Peningkatan Hasil Diagnosis Kanker Payudara*.

Rubio, I. T., Diaz-Botero, S., Esgueva, A., … Espinosa-Bravo, M. (2015). The superparamagnetic iron oxide is equivalent to the Tc99 radiotracer method for identifying the sentinel lymph node in breast cancer. *European Journal of Surgical Oncology*, *41*(1), 46–51.

https://doi.org/10.1016/j.ejso.2014.11.006

Syifa, R. A., Adi, K., & Widodo, C. E. (2016). Analisis Tekstur Citra Mikroskopis Kanker Paru Menggunakan Metode Gray Level Co-Occurance Matrix (Glcm) Dan Tranformasi Wavelet Dengan Klasifikasi Naive Bayes. *Youngster Physics Journal*, *5*(4), 457–462.

Ullah, Z., Khan, M. N., Din, Z. U., & Afaq, S. (2021). Breast Cancer Awareness and Associated Factors Amongst Women in Peshawar, Pakistan: A Cross-Sectional Study. *Breast Cancer: Basic and Clinical Research*, *15*. https://doi.org/10.1177/11782234211025346