

## PERBAIKAN KONTRAS CITRA MAMMOGRAM PADA KLASIFIKASI KANKER PAYUDARA BERDASARKAN FITUR GRAY-LEVEL CO-OCCURRENCE MATRIX

Febri Liantoni<sup>1</sup>, Agus Santoso<sup>2</sup>

<sup>1</sup>Pendidikan Teknik Informatika dan Komputer, Universitas Sebelas Maret  
Jl. Ir. Sutami No.36 A, Pucangsawit, Kec. Jebres, Kota Surakarta, Jawa Tengah

<sup>2</sup>Teknik Informatika, Institut Teknologi Adhi Tama Surabaya  
Jl. Arief Rachman Hakim 100, Klampis Ngasem, Sukolilo, Surabaya, Jawa Timur

e-mail: [febri.liantoni@gmail.com](mailto:febri.liantoni@gmail.com)<sup>1</sup>, [agus.santoso@gmail.com](mailto:agus.santoso@gmail.com)<sup>2</sup>

Received : February, 2020

Accepted : April, 2020

Published : April, 2020

### Abstract

*In this era to recognize breast tumors can be based on mammogram images. This method will expedite the process of recognition and classification of breast cancer. This research was conducted classification techniques of breast cancer using mammogram images. The proposed model targets classification studies for cases of malignant, and benign cancer. The research consisted of five main stages, preprocessing, histogram equalization, convolution, feature extraction, and classification. For preprocessing cropping the image using region of interest (ROI), for convolution, median filter and histogram equalization are used to improve image quality. Feature extraction using Gray-Level Co-Occurrence Matrix (GLCM) with 5 features, entropy, correlation, contrast, homogeneity, and variance. The final step is the classification using Radial Basis Function Neural Network (RBFNN) and Support Vector Machine (SVM). Based on the hypotheses that have been tested and discussed, the accuracy for RBFNN is 86.27%, while the accuracy for SVM is 84.31%. This shows that the RBFNN method is better than SVM in distinguishing types of breast cancer. These results prove the process of improving image construction using histogram equalization and the median filter is useful in the classification process.*

**Keywords:** gray-level co-occurrence matrix, mammogram, radial basis function neural network, region of interest, support vector machine.

### Abstrak

*Pada era sekarang ini untuk mengenali tumor payudara bisa berdasarkan citra mammogram. Dengan cara ini akan mempermudah proses pengenalan dan pengklasifikasian kanker payudara. Pada penelitian ini dilakukan sebuah teknik klasifikasi kanker payudara dengan menggunakan citra mammogram. Model yang diusulkan memiliki target klasifikasi untuk kasus kanker ganas dan kanker jinak. Penelitian yang dilakukan terdiri dari lima tahap utama yaitu preprosesing, ekualisasi histogram, konvolusi, ekstraksi fitur dan klasifikasi. Untuk preprosesing dilakukan cropping citra menggunakan region of interest (ROI), untuk konvolusi digunakan median filter dan ekualisasi histogram yang bertujuan untuk peningkatan kualitas citra. Ekstraksi fitur menggunakan Gray-Level Co-Occurrence Matrix (GLCM) dengan 5 fitur yaitu entropy, correlation, contrast, homogeneity dan variance. Langkah terakhir adalah klasifikasi menggunakan Radial Basis Function Neural Network (RBFNN) dan Support Vector Machine (SVM). Berdasarkan hipotesis yang telah diuji dan dibahas diperoleh akurasi untuk RBFNN sebesar 86.27%, sedangkan akurasi untuk SVM sebesar 84.31%. Hal ini menunjukkan bahwa metode RBFNN lebih baik dibandingkan SVM dalam membedakan jenis kanker payudara. Hasil ini membuktikan proses peningkatan kontras citra menggunakan ekualisasi histogram dan median filter bermanfaat dalam proses klasifikasi.*

**Kata Kunci:** *gray-level co-occurrence matrix, mammogram, radial basis function neural network, region of interest, support vector machine.*

## 1. PENDAHULUAN

Mammografi adalah metode yang efektif untuk pemeriksaan kanker payudara. Faktor resiko yang menyebabkan penyakit kanker payudara antara lain jenis kelamin, obesitas, usia, riwayat reproduksi, riwayat keluarga, dan konsumsi makanan berlemak tinggi [1]. Meskipun perawatan kanker payudara, termasuk pembedahan, radioterapi, kemoterapi, dan terapi hormon telah banyak dilakukan namun teknik-teknik ini menyebabkan pasien berpotensi menderita efek buruk yang substansial bahkan kematian. Pengobatan dengan cara pembedahan dan kemoterapi dapat dilakukan tetapi belum tentu berhasil jika kanker telah mencapai stadium akhir [2]. Tes dengan pencitraan dapat dilakukan untuk mengetahui area mencurigakan akan menjadi kanker. Tes pencitraan ini seperti dengan memanfaatkan teknologi *ultrasound*, *mammogram*, dan *magnetic resonance imaging* (MRI) [3].

Di era teknologi saat ini, sistem diagnostik banyak menggunakan metode skrining mammogram untuk mengklasifikasikan tumor payudara. Sistem *Computer Aided Diagnosis* (CAD) biasanya mengandalkan teknik pembelajaran mesin untuk mendeteksi kanker dalam gambar mammogram digital [3]–[6]. Teknik seperti itu dilakukan dengan fitur diskriminatif dan deskriptif untuk mengklasifikasikan citra menjadi beberapa kelas. Dalam dekade terakhir banyak metode telah diusulkan untuk mengklasifikasikan citra mammogram dan untuk mendapatkan akurasi, efisiensi, ketahanan, dan presisi yang lebih baik. Namun demikian itu masih merupakan area penelitian terbuka karena tantangan intrinsik dalam representasi dan klasifikasi mammogram. Banyak peneliti telah mempelajari gambar mammogram untuk klasifikasi dua kelas (normal dan abnormal) dan mencapai hasil yang signifikan. Mazurowski et al. mengusulkan sebuah template berdasarkan pada algoritma pengenalan untuk massa payudara [7]. Kumpulan data mereka didasarkan pada data Digital Database untuk Screening Mammography (DDSM) gambar dan mencapai akurasi hingga 83%. Klasifikasi kanker payudara

berdasarkan citra mammogram juga dilakukan dengan membandingkan kinerja dukungan berbasis mesin vektor (SVM) dengan algoritma tetangga terdekat [8], [9]. Listia melakukan klasifikasi citra mammogram berdasarkan GLCM dengan metode Backpropagation dengan rata-rata akurasi sebesar 81,1%. Pada penelitian ini digunakan mean filter untuk preproses citra mammogram [10]. Penggunaan ekualisasi histogram pernah digunakan peneliti pada penelitian sebelumnya mengenai klasifikasi jenis telur biasa dan telur omega-3 berdasarkan ciri orde pertama [11].

Peningkatan kontras citra mammogram dengan memanfaatkan ekualisasi histogram dan median filter menjadi salah satu tujuan pada penelitian ini. Proses ini digunakan untuk meningkatkan nilai ekstraksi fitur pada GLCM. Metode RBFNN dan SVM digunakan untuk mengklasifikasikan kanker payudara berdasarkan citra mammogram dari fitur GLCM. Citra *mammograph image analysis society* (MIAS) digunakan sebagai data acuan yang diolah. Penelitian ini diharapkan dapat bermanfaat dan berkontribusi untuk pengklasifikasian kanker payudara ganas dan kanker jinak dengan lebih cepat.

## 2. METODE PENELITIAN

Beberapa tahapan koleksi dataset citra mammogram, proses ekstraksi region of interest (ROI), peningkatan citra menggunakan ekualisasi histogram dan median filter, ekstraksi fitur dengan GLCM dan klasifikasi menggunakan RBFNN dan SVM.

### 2.1 Dataset citra mammogram

Dataset mamogram diambil dari database MIAS yang dipakai sebagai dataset penelitian [12]. Format data citra yang digunakan berekstensi \*.pgm.

### 2.2 Region Of Interest (ROI)

Pada penelitian ini citra mammogram dilakukan cropping untuk mendeteksi massa secara proporsional menggunakan region of interest. Pada dataset citra mammogram yang digunakan masih berbentuk tak beraturan dan belum ternormalisasi. ROI digunakan untuk mengambil

bagian yang akan diteliti sehingga dapat mengurangi kesalahan dalam proses klasifikasi.

### 2.3 Ekualisasi Histogram

Ekualisasi histogram merupakan proses pemerataan histogram, dimana distribusi nilai derajat keabuan pada suatu citra dibuat rata [5]. Untuk dapat melakukan ekualisasi histogram diperlukan suatu fungsi distribusi kumulatif. Fungsi distribusi kumulatif seperti didefinisikan berikut ini.

$$f(k) = \frac{(N-1)}{M} \sum_{k=0}^n h(k); \quad (1)$$

Dimana M merupakan nilai keseluruhan pixel dalam suatu citra. N merupakan nilai piksel keabuan. Dan  $h(k)$  merupakan histogram pada nilai gray value k.

### 2.4 Ekstraksi fitur GLCM

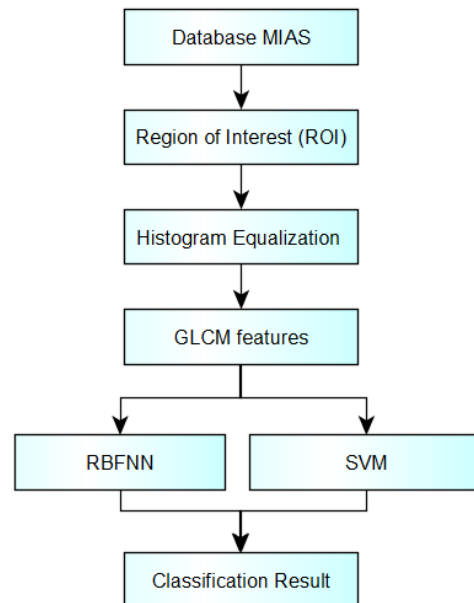
*Gray-Level Co-Occurrence Matrix* (GLCM) digunakan karena metode ini bekerja untuk data citra gray. GLCM pada dasarnya adalah distribusi probabilitas gabungan tingkat abu-abu pada pasangan posisi yang memuatkan posisi relatif tertentu dalam suatu citra [13]. GLCM yang digunakan sebanyak 5 fitur antara lain fitur *entropy, correlation, contrast, homogeneity dan variance*. Fitur ini sering digunakan pada beberapa penelitian dan terbukti menghasilkan data yang bagus [14].

### 2.5 Klasifikasi

*Radial Basis Function* (RBF) merupakan bagian dari metode *Artificial Neural Network* (ANN) berdasarkan fungsi radial basis sebagai fungsi aktivasi [15], [16]. Arsitektur RBF terdapat tiga lapisan yaitu lapisan input, tersembunyi, dan output. RBFNN memiliki fungsi sigmoid dalam fungsi aktivasinya. Ciri khas RBFNN ini menjadikan RBFNN memiliki proses kerja lebih cepat dibandingkan dengan algoritme NN lainnya [17].

*Support Vector Machine* (SVM) merupakan teori pembelajaran statistik dan dapat memberikan hasil yang lebih baik dari pada metode yang lain. SVM dapat bekerja dengan baik pada data dengan berdimensi set tinggi [8]. SVM menggunakan teknik kernel dan hanya sejumlah data yang terpilih yang berkontribusi untuk membangun model klasifikasi. Hal tersebut menjadi kelebihan SVM, karena tidak semua data latih akan dilihat untuk dilibatkan dalam setiap iterasi pelatihannya.

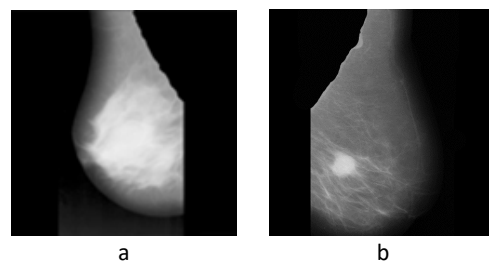
Secara keseluruhan tahapan dari penelitian ini seperti ditunjukkan pada Gambar 1 berikut ini.



Gambar 1. Tahapan penelitian

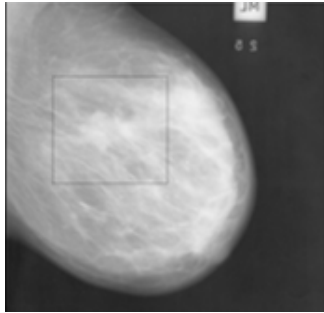
## 3. HASIL DAN PEMBAHASAN

Proses uji coba dilakukan untuk mengetahui tingkat keberhasilan sistem dalam mengklasifikasikan kanker ganas dan kanker jinak. Proses pengujian dengan 111 citra mammogram terdiri dari 62 citra kanker jinak dan 49 citra kanker ganas. Data tersebut kemudian dibagi menjadi data pelatihan sebanyak 60 citra dan data uji sebanyak 51 citra. Contoh data *mammogram* seperti ditunjukkan pada Gambar 2.



Gambar 2. Citra mammogram, a) kanker jinak, b) kanker ganas

Tahap awal penelitian dengan melakukan proses *cropping* menggunakan ROI. Proses ini dilakukan untuk mengambil bagian yang akan diteliti pada penelitian ini. Proses ROI ditunjukkan pada Gambar 3.



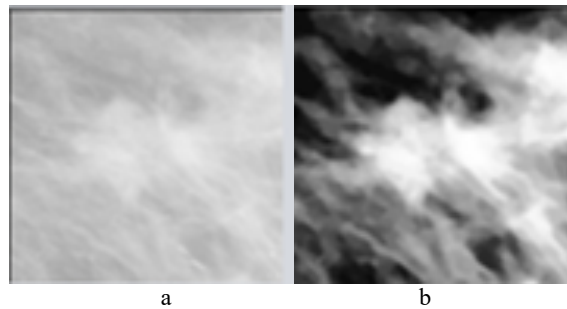
Gambar 3. Proses ROI

Proses selanjutnya dengan dilakukan histogram ekualisasi dan median filter. Proses ini bertujuan untuk memperbaiki kualitas citra dengan cara mengubah sebaran tingkat keabuan citra sehingga diharapkan memperoleh nilai fitur yang optimal. Dari kedua proses ini peningkatan kontras citra dan penghilangan noise yang mungkin muncul akibat pengolahan citra mammogram dilakukan. Proses peningkatan

kontras citra merupakan salah satu usulan yang dilakukan pada penelitian ini.

Setelah proses peningkatan kontras citra selanjutnya dilakukan ekstraksi fitur menggunakan GLCM. Proses klasifikasi kanker payudara menggunakan citra mammogram dilakukan dengan berdasarkan fitur *entropy*, *correlation*, *contrast*, *homogeneity* dan *variance*. Contoh hasil ekstraksi fitur GLCM seperti Tabel 1.

Proses pengujian klasifikasi dilakukan dengan membandingkan metode RBFNN dan SVM. Uji coba dilakukan untuk 51 citra data uji terdiri dari 22 citra kanker jinak dan 29 citra kanker ganas. Proses pengujian dengan menggunakan cara *precision*, *recall* dan *accuracy* dinilai cocok untuk kasus klasifikasi [18]. Hasil pengujian dari matriks confusion seperti ditunjukkan pada Tabel 2, Tabel 3, dan Tabel 4.



Gambar 4. Peningkatan kontras, a) sebelum, b) setelah

Tabel 1. Contoh hasil ekstraksi GLCM

Filename	Correlation	Contrast	Homogeneity	Variance	Entropy
mdb195.pgm	16.75808724546655	1.1680070157015316	0.861901824158197	50.83799503204632	1.5013852299892925
mdb198.pgm	15.809768473279266	0.4518950483213467	0.8808060339684876	50.85548425242172	1.4816096895554471
mdb199.pgm	17.29396528133651	0.4594124422623845	0.8667503765454783	53.55148497380021	1.4973713869374674
mdb204.pgm	18.02743188409681	0.5758630853199207	0.8596189198480847	57.386995484367304	1.3633715241103017
mdb207.pgm	19.758066767787938	0.5727588408301816	0.8749344391575438	59.925914984508694	1.4171337407936295

Tabel 2. Model matriks confusion algoritme RBFNN

Kelas Prediksi	True Ya	True Tidak	Class Precision
Prediksi Ya	26	3	89.65%
Prediksi Tidak	4	18	81.81%
Class Recall	86.66%	85.71%	

Tabel 3. Model matriks confusion algoritme SVM

Kelas Prediksi	True Ya	True Tidak	Class Precision
Prediksi Ya	24	3	88.88%
Prediksi Tidak	5	19	79.16%
Class Recall	82.75%	86.36%	

Tabel 4. Hasil komparasi algoritme RBFNN dan SVM

Parameter	RBFNN	SVM
Accuracy	86.27%	84.31%

Matriks confusion dari algoritme RBFNN pada Tabel 2 menunjukkan bahwa dari 51 data uji menghasilkan nilai presisi sebesar 89.65% untuk prediksi Ya dan 81.81% untuk prediksi Tidak. Sedangkan nilai recall sebesar 86.66% untuk Ya dan 85.71% untuk Tidak. Akurasi untuk RBFNN yang diperoleh sebesar 86.27%.

Matriks confusion dari algoritme SVM pada Tabel 3 menunjukkan bahwa dari 51 data uji menghasilkan nilai presisi sebesar 88.88% untuk prediksi Ya dan 79.16% untuk prediksi Tidak. Sedangkan nilai recall sebesar 82.75% untuk Ya dan 86.36% untuk Tidak. Akurasi untuk SVM yang diperoleh sebesar 84.31%.

Hasil akurasi yang diperoleh sebanding dengan penelitian oleh Mazurowski sebesar 83%, walaupun menggunakan data uji yang berbeda [7]. Hasil penelitian juga sesuai dengan penelitian listia dengan fitur GLCM dan metode klasifikasi backpropagation rata-rata akurasi sebesar 81,1% [10]. Hasil pengujian berbeda juga disebabkan pada penelitian ini tidak menggunakan data citra untuk kelas normal.

#### 4. KESIMPULAN

Berdasarkan hipotesis yang telah diuji dan dibahas diperoleh kesimpulan bahwa proses peningkatan kontras yang dilakukan pada sistem mampu mengklasifikasi kanker ganas dan kanker jinak. Hasil pengujian menunjukkan bahwa RBFNN memperoleh akurasi sebesar 86.27% sedangkan SVM memperoleh akurasi sebesar 84.31%. Hal ini menunjukkan algoritme RBFNN lebih baik dibandingkan algoritme SVM. Hasil ini juga menunjukkan bahwa RBFNN dan SVM berdasarkan fitur GLCM mampu dijadikan acuan untuk membedakan jenis kanker payudara. Berdasarkan hasil pengujian tersebut juga membuktikan proses peningkatan kontras citra menggunakan ekualisasi histogram dan

median filter bermanfaat dalam proses klasifikasi.

#### DAFTAR PUSTAKA

- [1] Z. Kleibl and V. N. Kristensen, "Women at high risk of breast cancer: Molecular characteristics, clinical presentation and management," *The Breast*, vol. 28, pp. 136–144, Aug. 2016.
- [2] A. Wanchai, J. M. Armer, B. R. Stewart, and B. B. Lasinski, "Breast cancer-related lymphedema: A literature review for clinical practice," *Int. J. Nurs. Sci.*, vol. 3, no. 2, pp. 202–207, Jun. 2016.
- [3] D. Daye *et al.*, "Mammographic Parenchymal Patterns as an Imaging Marker of Endogenous Hormonal Exposure," *Acad. Radiol.*, vol. 20, no. 5, pp. 635–646, May 2013.
- [4] J. Chaki, R. Parekh, and S. Bhattacharya, "Plant leaf classification using multiple descriptors: A hierarchical approach," *J. King Saud Univ. - Comput. Inf. Sci.*, 2018.
- [5] M. M. Jadoon, Q. Zhang, I. U. Haq, S. Butt, and A. Jadoon, "Three-Class Mammogram Classification Based on Descriptive CNN Features," *Biomed Res. Int.*, vol. 2017, pp. 1–11, 2017.
- [6] D. Pandey *et al.*, "Automatic and fast segmentation of breast region-of-interest (ROI) and density in MRIs," *Heliyon*, vol. 4, no. 12, p. e01042, Dec. 2018.
- [7] M. A. Mazurowski, J. Y. Lo, B. P. Harrawood, and G. D. Tourassi, "Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis," *J. Biomed. Inform.*, vol. 44, no. 5, pp. 815–823, Oct. 2011.
- [8] K. Novar Setiawan and I. M. Suwija Putra, "Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM)," *J. Ilm.*

- Merpati (Menara Penelit. Akad. Teknol. Informasi)*, p. 13, Apr. 2018.
- [9] F. Shirazi and E. Rashedi, "Detection of cancer tumors in mammography images using support vector machine and mixed gravitational search algorithm," in *2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*, 2016, pp. 98–101.
- [10] R. Listia and A. Harjoko, "Klasifikasi Massa pada Citra Mammogram Berdasarkan Gray Level Cooccurrence Matrix (GLCM)," *IJCCS*, vol. 8, no. 1, 2014.
- [11] F. Liantoni and A. A. Santoso, "Penerapan Ekstraksi Ciri Statistik Orde Pertama Dengan Ekualisasi Histogram Pada Klasifikasi Telur Omega-3," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 953–958, 2018.
- [12] J. Parker *et al.*, "Mammographic Image Analysis Society (MIAS) database v1.21," 2015.
- [13] G. Zheng *et al.*, "Development of a Gray-Level Co-Occurrence Matrix-Based Texture Orientation Estimation Method and Its Application in Sea Surface Wind Direction Retrieval From SAR Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5244–5260, Sep. 2018.
- [14] S. Suharjito, B. Imran, and A. S. Girsang, "Family Relationship Identification by Using Extract Feature of Gray Level Co-occurrence Matrix (GLCM) Based on Parents and Children Fingerprint," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 5, p. 2738, Oct. 2017.
- [15] P. Y. Goh, S. C. Tan, W. P. Cheah, and C. P. Lim, "Adaptive rough radial basis function neural network with prototype outlier removal," *Inf. Sci. (Ny.)*, vol. 505, pp. 127–143, Dec. 2019.
- [16] C. Jayaweera, M. Othman, and N. Aziz, "Improved predictive capability of coagulation process by extreme learning machine with radial basis function," *J. Water Process Eng.*, vol. 32, p. 100977, Dec. 2019.
- [17] A. Hashemi Fath, F. Madanifar, and M. Abbasi, "Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, Dec. 2018.
- [18] A. Fadli, M. I. Zulfa, and Y. Ramadhani, "Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 4, p. 158, Oct. 2018.