

Klasifikasi Tumor Payudara Pada Citra Ultrasonografi Menggunakan Multi-fitur Tekstur dan *Support Vector Machine*

Rahmad Hidayat^{1*}, Huzaeni¹, Mahdi¹, M. Khadafi¹, Muhammad Davi¹

¹Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe

*Koresponden email: rahmad_hidayat@pnl.ac.id

Diterima: 13 Februari 2024

Disetujui: 9 April 2024

Abstract

Breast cancer is a prevalent type of cancer affecting women worldwide. Additionally, Globocan reported nearly 400,000 new cancer cases in Indonesia in 2020, with 16% being breast cancer. The Ministry of Health has prioritised breast cancer treatment due to the high number of cases. Early detection is a crucial factor in increasing patient life expectancy. Stage 1 breast cancer, for instance, has a 5-year life expectancy of 100%. Breast ultrasound or mammary ultrasound is a commonly used method to detect various breast problems, including cysts and tumors. It is a relatively easy procedure, and the necessary equipment is generally available at Health Facility 1. Texture features are extracted from breast ultrasound images using Gabor and Gray Level Co-occurrence Matrix (GLCM) techniques. The resulting feature vector is then selected and its dimensions reduced to simplify the computing process. This vector is then used to train an SVM classifier to distinguish between benign and malignant cases. The accuracy of the classifier is 0.67 (training) and 0.66 (validation). Meanwhile, the loss obtained during training was 0.77 and during validation was 0.84. Further improvement is required for the accuracy of the model to be applicable.

Keywords: *breast tumor, features extraction, texture, classification*

Abstrak

Kanker payudara merupakan salah satu jenis kanker yang banyak mengancam perempuan di berbagai belahan dunia. Selain itu, data Globocan juga menunjukkan bahwa pada tahun 2020 terdapat hampir 400 ribu kasus kanker baru di Indonesia dengan 16%-nya merupakan kanker payudara. Oleh karena tingginya kasus kanker payudara, pemerintah melalui kementerian kesehatan menjadikan kanker payudara sebagai prioritas penanganan. Salah satu faktor yang dapat meningkatkan harapan hidup pasien adalah deteksi dini, dimana kanker payudara stadium 1 memiliki angka harapan hidup 5 tahun adalah 100%. Salah satu metode yang sering digunakan untuk mendeteksi berbagai masalah pada payudara (seperti kista dan tumor) adalah USG payudara atau USG mammae. USG payudara relatif mudah dan alatnya umumnya tersedia pada Fasilitas 1. Penelitian ini melakukan ekstraksi fitur tekstur yaitu Gabor dan GLCM (*Gray Level Co-occurrence Matrix*) dari citra USG payudara. Vektor fitur yang terbentuk selanjutnya diseleksi dan direduksi dimensinya agar meringankan proses komputasi. Vektor fitur tersebut kemudian selanjutnya digunakan untuk melatih klasifikasi SVM untuk mengklasifikasi kedalam dua kelas yaitu benign (tumor jinak) dan malignant (tumor ganas/kanker Berdasarkan hasil pengujian didapatkan akurasi sebesar 0.67 (training) dan 0.66 (validasi). sementara itu, loss yang didapat adalah sebesar 0.77 saat training dan 0.84 saat validasi. Akurasi yang didapatkan masih perlu ditingkatkan jika model yang dikembangkan ingin diterapkan pada level aplikatif.

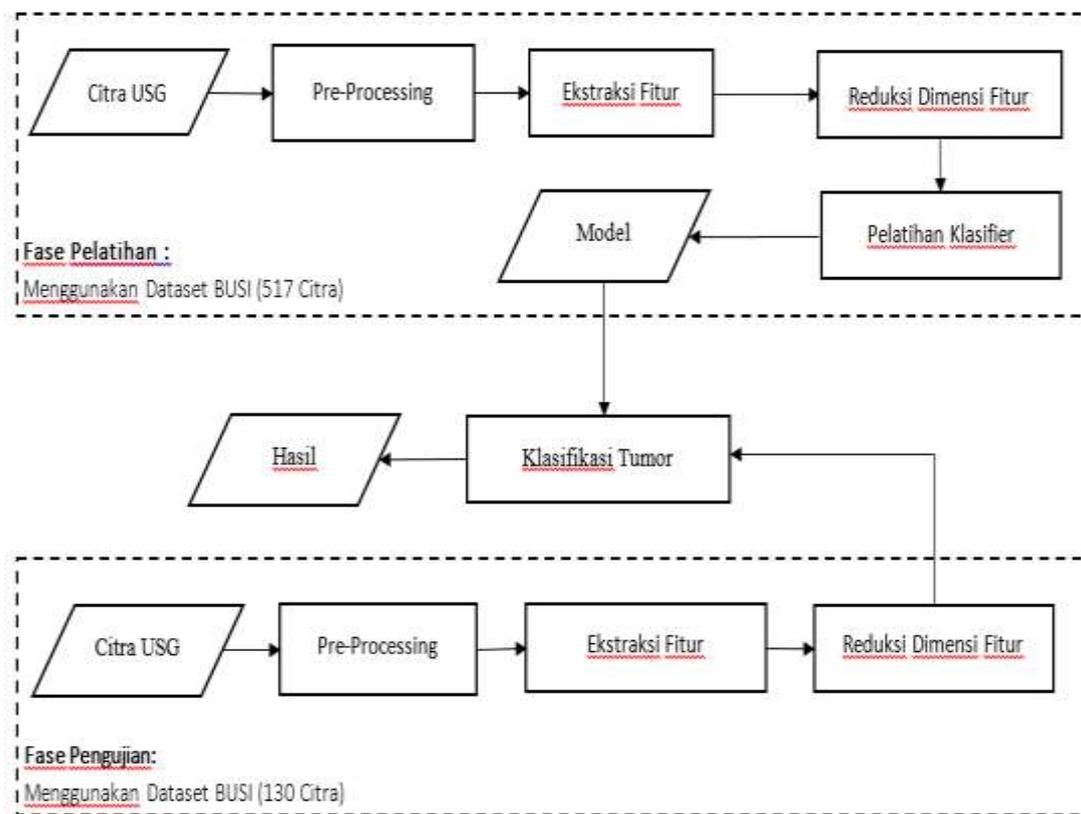
Keywords: *tumor payudara, ekstraksi fitur, tekstur, klasifikasi*

1. Pendahuluan

Kanker payudara adalah suatu jenis tumor ganas yang berkembang pada sel-sel payudara. Kanker ini dapat tumbuh jika terjadi pertumbuhan yang abnormal dari sel-sel pada payudara. Sel-sel tersebut membelah diri lebih cepat dari sel normal dan berakumulasi, yang kemudian membentuk benjolan atau massa. Berdasarkan data Globocan tahun 2020, jumlah kasus baru kanker payudara mencapai 68.858 kasus (16,6%) dari total 396.914 kasus baru kanker di Indonesia [1]. Menurut data terakhir WHO, angka kematian karena kanker payudara di Indonesia mencapai 1,41% dari seluruh kematian atau angka kematian disesuaikan dengan usia adalah 2.025 per 100.000 penduduk [2]. Tingginya angka kanker payudara di Indonesia menjadi prioritas penanganan oleh pemerintah. Tumor payudara ini secara umum dibagi menjadi 2, yaitu benign (tumor jinak) dan malignant (tumor ganas). Tumor payudara jinak ditandai dengan berbentuk benjolan kecil bulat dan lembut. Tumor payudara dalam tingkat jinak biasanya akan mempunyai keadaan dan pertumbuhan yang tidak bersifat kanker. Kanker ini bisa terdeteksi tetapi tidak akan menular

dan merusak jaringan di dekatnya. Pada tumor payudara dalam tingkat ganas ditandai dengan bentuk yang tidak simetris, kasar, terasa nyeri, dan lainnya. Deteksi dini merupakan salah satu cara untuk menekan angka kematian selain menghindari faktor resiko lainnya. USG payudara atau USG mammae adalah salah satu jenis USG yang secara khusus dilakukan untuk memeriksa kondisi payudara dan mendeteksi gangguan serta berbagai bentuk kelainan pada payudara, seperti kista dan tumor. USG payudara relatif mudah dilakukan dan peralatannya banyak tersedia di Faskes 1.

Penelitian ini melakukan ekstraksi fitur tekstur yaitu Gabor dan GLCM (*Gray Level Co-occurrence Matrix*) dari citra USG payudara. Gabor merupakan kumpulan filter yang digunakan untuk mengekstrak fitur tekstur dari citra. Gabor filter merupakan fungsi gaussian yang merupakan hasil modulasi frekuensi dan orientasi sinusoid kompleks. Gaussian filter memiliki kemampuan ekstraksi baik secara spasial maupun domain pada berbagai dimensi. Penelitian yang dilakukan oleh [3] memanfaatkan fitur ini untuk deteksi kanker payudara pada citra Thermogram. Model yang dibangun dengan menggunakan metode SVM menunjukkan akurasi yang kompetitif yaitu sebesar 84,5%. *Gray-Level Co-occurrence matrix* (GLCM) merupakan teknik yang perhitungan statistik pada citra dan juga sering digunakan untuk menganalisis tekstur pada citra. GLCM merepresentasikan hubungan antara 2 pixel yang bertetangga (*neighboring pixels*) yang memiliki intensitas keabuan (*grayscale intensity*), jarak dan sudut. Terdapat 8 sudut yang dapat digunakan pada GLCM, diantaranya sudut 0°, 45°, 90°, 135°, 180°, 225°, 270°, atau 315°. Deteksi kanker dengan memanfaatkan fitur ini telah banyak dilakukan antara lain pada [2], [4]–[6].



Gambar 1. Tahapan Model Klasifikasi Tumor Payudara

Hasil pengujian menunjukkan fitur ini cukup efektif terutama jika pada dataset ukuran kecil, namun pada dataset berukuran besar perlu dikombinasikan dengan fitur lainnya. Vektor fitur yang terbentuk selanjutnya direduksi dengan menggunakan metode PCA (*Principal Component Analysis*). Setelah direduksi, vektor fitur selanjutnya digunakan untuk melatih klasifier SVM untuk mengklasifikasi kedalam dua kelas yaitu benign (tumor jinak) dan malignant (tumor ganas/kanker). Karena data yang akan diklasifikasi kompleks dan tidak linear maka SVM dilatih dengan menggunakan kernel. Adapun kernel yang akan digunakan ditentukan setelah diuji tingkat validitas klasifikasinya. Setelah proses pelatihan selesai dilakukan maka model yang dihasilkan kemudian diuji dengan menggunakan data testing. Performa klasifikasi model kemudian diukur dengan menggunakan confusion matrix.

2. Metode Penelitian

Secara umum metode yang dikembangkan beberapa tahapan antara lain pre-processing, ekstraksi fitur, pelatihan dan pengujian. Fase pelatihan terdiri dari proses pre-processing, ekstraksi fitur Gabor, ekstraksi fitur GLCM, reduksi dimensi vektor fitur, dan pelatihan klasifier. Pada fase pengujian terdiri dari pre-processing, ekstraksi fitur Gabor dan GLCM, dan klasifikasi. **Gambar 1** menampilkan tahapan yang terdapat pada penelitian ini dengan lebih detail.

a. Pre-processing

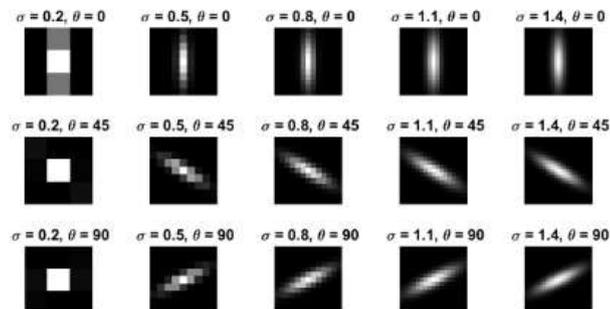
Pada sebagian besar sistem Computer Vision, *smoothing* merupakan langkah yang biasa digunakan pada tahap pre-processing untuk menghilangkan *noise*. *Smoothing* juga dapat digunakan untuk menghilangkan *noise* yang ditimbulkan oleh alam seperti hujan dan salju yang direkam oleh kamera yang diletakan diluar ruangan [12]. Selain *smoothing*, tahapan lain yang juga dapat meningkatkan kualitas video adalah perbaikan kontras yang bertujuan untuk memberikan distribusi kontras dengan lebih baik.

b. Ekstraksi fitur

Gabor merupakan kumpulan filter yang digunakan untuk mengekstrak fitur tekstur dari citra [13]. Gabor filter merupakan fungsi gaussian yang merupakan hasil modulasi frekuensi dan orientasi sinusoid kompleks. Gaussian filter memiliki kemampuan ekstraksi baik secara spasial maupun domain pada berbagai dimensi [14]. Pada domain spasial, kernel Gabor filter 2D merupakan kernel Gaussian yang merupakan hasil modulasi dari gelombang sinusoidal seperti yang ditampilkan pada persamaan dibawah ini.

$$G(x, y, \theta, \sigma, \lambda, \psi, \gamma) = \frac{\exp(-(x'^2 + y'^2/\gamma^2))}{2\sigma^2} \times \exp(i(\frac{2\pi x'}{\lambda} + \psi))$$

Dimana $\sigma, \psi, \lambda, \gamma$ merupakan ukuran filter, pergerakan fasa, panjang gelombang dan aspek rasio spasial sinusoid secara berurutan. x' dan y' dikalkulasi dari orientasi spasial, θ merupakan dari persamaan. Himpunan filter Gabor diekstraksi dengan ukuran filter σ dan orientasi θ . **Gambar 2** menampilkan visualisasi filter Gabor pada berbagai orientasi dan ukuran.



Gambar 2. Visualisasi filter Gabor pada berbagai orientasi dan ukuran

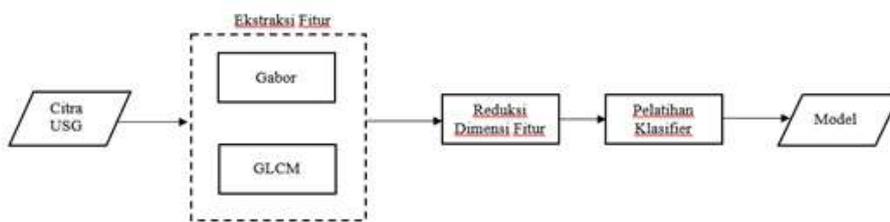
Matrik GLCM dihitung dari nilai piksel yang berpasangan dan memiliki nilai intensitas tertentu [15]. Jika jarak antara dua piksel yaitu (x_1, y_1) dan (x_2, y_2) , dan Θ didefinisikan sebagai sudut antara kedua piksel tersebut, maka matriks GLCM merupakan distribusi spasial dari P_d, Θ . Ekstraksi fitur GLCM dilakukan pada arah sudut dengan jarak 1 piksel dan 4 jenis sudut yang digunakan yaitu 00 (1800), 450 (2250), 900 (2700), 1350 (3150). Jika i adalah citra yang terdapat dalam database DB, maka notasi untuk masing-masing fitur tekstur pada keempat sudut tersebut dapat dilihat pada **Tabel 1**.

Tabel 1. Notasi fitur GLCM

No.	Ruang Warna	Sudut			
		0°	45°	90°	135°
1.	Energi	g_0^{eng}	g_1^{eng}	g_2^{eng}	g_3^{eng}
2.	Entropi	g_0^{ent}	g_1^{ent}	g_2^{ent}	g_3^{ent}
3.	Kontras	g_0^{hom}	g_1^{hom}	g_2^{hom}	g_3^{hom}
4.	Korelasi	g_0^{con}	g_1^{con}	g_2^{con}	g_3^{con}
5.	Homogenitas	g_0^{cor}	g_1^{cor}	g_2^{cor}	g_3^{cor}

c. Reduksi Dimensi Fitur

Metode analisis faktor bertujuan untuk mereduksi variabel dengan menghilangkan faktor-faktor dalam variabel tersebut yang saling terkait satu dan lainnya [16]. Salah satu metode yang cukup sering digunakan untuk mereduksi dimensi fitur adalah PCA. **Gambar 3** menampilkan tahapan reduksi dimensi vektor fitur dalam penelitian ini.

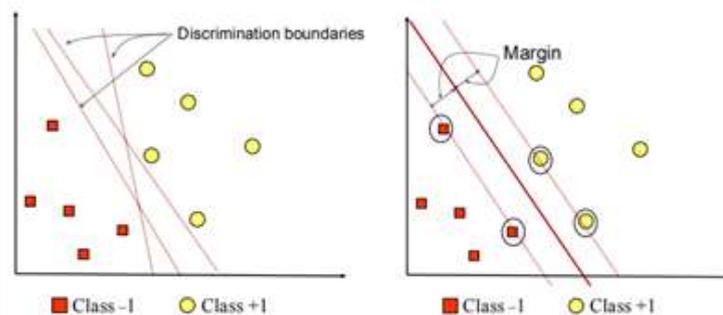


Gambar 3. Tahapan reduksi dimensi vektor fitur

Keunggulan metode analisis faktor diantaranya adalah dapat menghilangkan korelasi secara bersih tanpa harus mengurangi jumlah variabel asal. Langkah-langkah analisis faktor adalah sebagai berikut:

d. Support Vector Machine (SVM)

Setelah ekstraksi (menggunakan metode GLCM) dan reduksi dimensi fitur selesai dilakukan dengan, maka vektor fitur yang terbentuk selanjutnya akan digunakan untuk melatih model SVM. Model dilatih untuk dapat melakukan klasifikasi citra inputan kedalam dua kelas yaitu benign dan malignant. Terdapat beberapa kernel yang dapat digunakan untuk melatih model, namun kernel yang akan digunakan pada proses pengujian adalah kernel yang memberikan performa terbaik pada tahapan sebelumnya. Proses penentuan hyperplane untuk memisahkan dua buah kelas ditampilkan pada **Gambar 4** [17-20].



Gambar 4. Proses Penentuan Hyperplane Untuk Memisahkan Kelas -1 dan +1 (Liu dkk., 2008)

e. Dataset

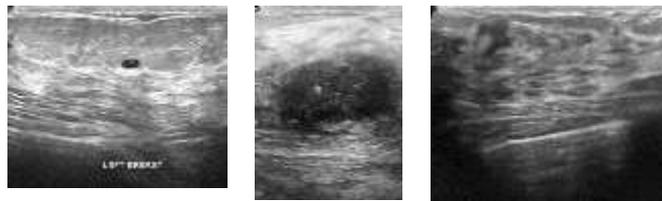
Penelitian ini menggunakan dataset BUSI untuk melatih dan menguji model klasifikasi tumor payudara pada citra USG. Dataset BUSI memiliki total citra sebanyak 647 dimana terdapat 437 citra tumor benign (jinak) dan 210 citra tumor tipe malignant (ganas). Fitur GLCM dan Gabor kemudian diekstraksi dari citra-citra tersebut, dimana vektor fitur yang terbentuk kemudian direduksi dan diseleksi dengan metode analisis komponen. menampilkan contoh citra pada dataset BUSI baik tumor yang bertipe benign maupun tumor yang bertipe malignant.

f. Skenario pengujian

Eksperimen pada penelitian ini dilakukan dengan menggunakan dua skenario. Pada skenario pertama (CA_SC1) pengujian dilakukan dengan menginputkan fitur GLCM saja, sementara pada skenario kedua (CA_SC2) fitur GLCM digabungkan secara linier dengan fitur gabor. Hasil pengujian untuk masing-masing skenario berupa kurva yang membandingkan *training* dan *validation* loss dengan menggunakan *threshold* yang berbeda-beda. Selain itu performa model juga diuji dengan membandingkan skor presisi, recall, dan F1 pada masing-masing skenario.

3. Hasil dan Pembahasan

Langkah *pre-processing* pada dataset dilakuan dengan memperbaiki kontras citra. Metode Histogram Equalization menggunakan komputasi yang sederhana dalam meningkatkan nilai kontras pada citra secara global, terutama jika nilai kontras pada citra terletak pada jarak yang berdekatan. Melalui penyesuaian ini, intensitas kontras dapat terdistribusi dengan lebih baik. Hasil citra sebelum dan setelah dilakukan preprocessing ditampilkan pada **Gambar 5**.



a. Citra sebelum diprocessing

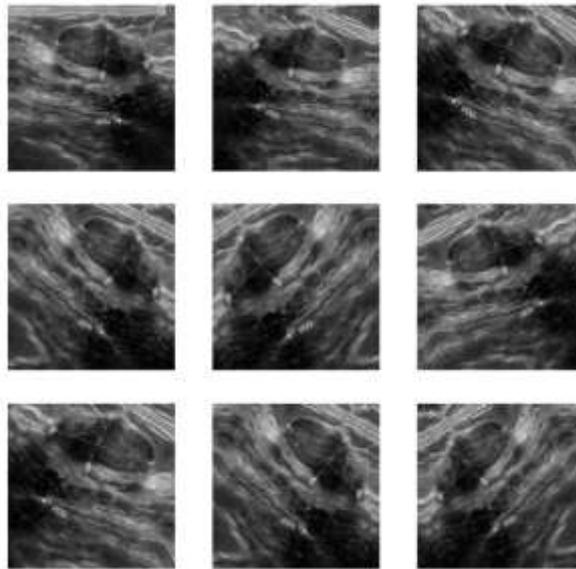


b. Citra setelah diprocessing

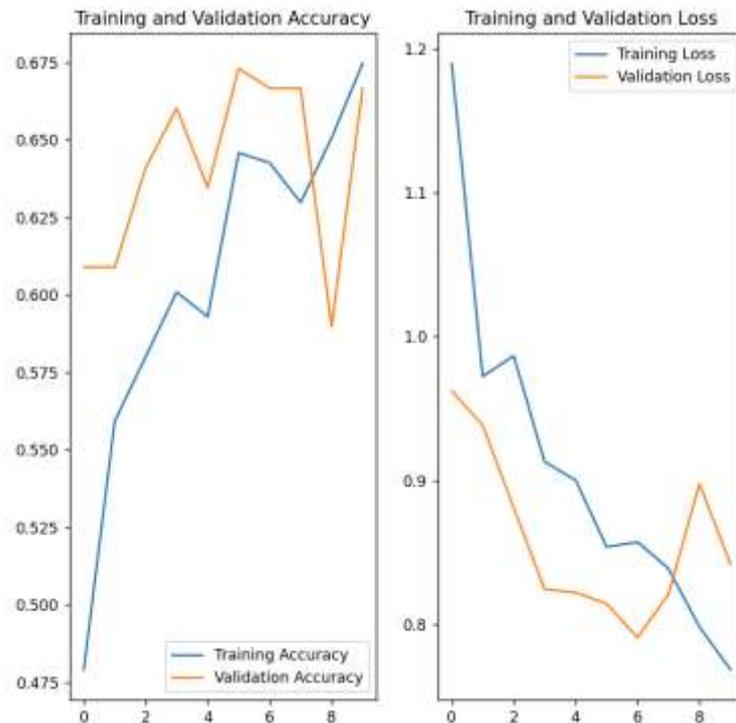
Gambar 5. Hasil *preprocessing* pada citra yang terdapat pada dataset BUSI

Untuk meningkatkan akurasi dan mencegah terjadinya *overfitting* (dimana model memasukan noise dan fitur-fitur yang tidak relevan sebagai bahan pembelajaran) maka dilakukan pendekatan dengan menggunakan data augmentasi. Data augmentasi membangkitkan data training tambahan dari data yang sudah ada dengan melakukan berbagai transformasi. Contoh citra hasil operasi augmentasi ditampilkan pada **Gambar 6**.

Setelah model selesai dilatih, selanjutnya dilakukan validasi terhadap performa model dalam melakukan klasifikasi. Pada penelitian ini sebanyak 80% data digunakan sebagai data latih dan sebanyak 20% digunakan sebagai data uji. Berdasarkan hasil pengujian didapatkan akurasi sebesar 0.67 (training) dan 0.66 (validasi). sementara itu, loss yang didapat adalah sebesar 0.77 saat training dan 0.84 saat validasi. Hasil pengujian model secara lebih detail ditampilkan pada **Gambar 7**.



Gambar 6. Citra hasil Augmentasi



Gambar 7. Hasil pengukuran Performa Model

4. Kesimpulan

Model klasifikasi tumor payudara yang dikembangkan pada penelitian ini telah menunjukkan akurasi yang cukup memuaskan. Berdasarkan hasil pengujian didapatkan akurasi sebesar 0.67 (training) dan 0.66 (validasi). sementara itu, loss yang didapat adalah sebesar 0.77 saat training dan 0.84 saat validasi. Akurasi yang didapatkan masih perlu ditingkatkan jika model yang dikembangkan ingin diterapkan pada level aplikatif. Pengembangan model klasifikasi tumor payudara pada tahapan selanjutnya dapat dilakukan dengan menambahkan tahapan antara lain: tahapan segmentasi dapat digunakan untuk melokalisasi area of interest (ROI). Penambahan tahapan segmentasi juga dapat menekan waktu pelatihan dan pengujian serta dapat meningkatkan akurasi. Pada citra-citra dengan kontras yang rendah harus terlebih dahulu diperbaiki. Penambahan fitur dan menemukan kombinasi fitur yang tepat pada tahapan klasifikasi juga dapat dilakukan untuk meningkatkan performa model yang dibangun.

5. Referensi

- [1] N. Micalef, D. Seychell, and C. J. Bajada, "Exploring the U-Net ++ Model for Automatic Brain Tumor Segmentation," *IEEE Access*, vol. 9, pp. 125523–125539, 2021, doi: 10.1109/ACCESS.2021.3111131.
- [2] M. Bella, A. Rasyid, F. Arnia, and K. Munadi, "Histogram Statistics and GLCM Features of Breast Thermograms for Early Cancer Detection," pp. 120–124, 2018.
- [3] A. A. Khan and A. S. Arora, "Breast Cancer Detection Through Gabor Filter Based Texture Features Using Thermograms Images," pp. 412–417, 2018.
- [4] T. Nguyen, T. Nguyen, and B. Ngo, "A GLCM Algorithm for Optimal Features of Mammographic Images for Detection of Breast Cancer," pp. 295–299, 2021.
- [5] M. Thohir, D. Candra, and R. Novitasari, "Classification of Colposcopy Data Using GLCM- SVM on Cervical Cancer," pp. 373–378, 2020.
- [6] T. T. Htay and S. S. Maung, "Early Stage Breast Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image," no. Iscit, pp. 171–175, 2018.
- [7] T. Imtiaz and S. Rifat, "Automated Brain Tumor Segmentation Based on Multi-Planar Superpixel Level Features Extracted From 3D MR Images," 2020.
- [8] A. S. Elkorany, M. Marey, K. M. Almustafa, and Z. F. Elsharkawy, "Breast Cancer Diagnosis Using Support Vector Machines Optimized by Whale Optimization and Dragonfly Algorithms," *IEEE Access*, vol. 10, no. July, pp. 69688–69699, 2022, doi: 10.1109/ACCESS.2022.3186021.
- [9] J. Jiang et al., "Multiple Resolution Residually Connected Feature Streams for Automatic Lung Tumor Segmentation From CT Images," vol. 38, no. 1, pp. 134–144, 2019.
- [10] C. Zhang and D. Wu, "A Stacked Multi-Connection Simple Reducing Net for Brain Tumor Segmentation," 2019.
- [11] B. Wang, X. Lv, M. I. N. Li, and J. Wang, "Classification of Cervical Biopsy Images Based on LASSO and EL-SVM," vol. 8, 2020.
- [12] T. Heseltine, N. Pears, and J. Austin, "Evaluation of image preprocessing techniques for eigenface-based face recognition," *Second Int. Conf. Image Graph.*, vol. 4875, no. 2002, p. 677, 2002, doi: 10.1117/12.477052.
- [13] D. Zhang, A. Wong, M. Indrawan, and G. Lu, "Content-based Image Retrieval Using Gabor Texture Features," *IEEE Trans. PAMI*, vol. 3656 LNCS, pp. 13–15, 2000.
- [14] A. N. Rumaksari, S. Sumpeno, and A. D. Wibawa, "Background subtraction using spatial mixture of Gaussian model with dynamic shadow filtering," *2017 Int. Semin. Intell. Technol. Its Appl. Strength. Link Between Univ. Res. Ind. to Support ASEAN Energy Sect. ISITIA 2017 - Proceeding*, vol. 2017-Janua, pp. 296–301, 2017, doi: 10.1109/ISITIA.2017.8124098.
- [15] S. Alameen and M. E. M. Gar-Elnabi, "Study of Glcm for Diagnosis of Liver Diseases From Abdominal Ct Images," no. August, 2016.
- [16] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010, doi: 10.1002/wics.101.
- [17] X. Y. Wang, L. L. Liang, W. Y. Li, D. M. Li, and H. Y. Yang, "A new SVM-based relevance feedback image retrieval using probabilistic feature and weighted kernel function," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 256–275, 2016, doi: 10.1016/j.jvcir.2016.03.008.
- [18] H. Yuan, "An SVM-Based AdaBoost Cascade Classifier for Sonar Image," vol. 8, pp. 115857–115864, 2020, doi: 10.1109/ACCESS.2020.3004473.
- [19] R. Liu, Y. Wang, T. Baba, D. Masumoto, and S. Nagata, "SVM-based active feedback in image retrieval using clustering and unlabeled data," *Pattern Recognit.*, vol. 41, no. 8, pp. 2645–2655, 2008, doi: 10.1016/j.patcog.2008.01.023.
- [20] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," *Proc. fifth Annu. Work. Comput. Learn. theory*, vol. 5, no. 3, pp. 144–152, 1992, doi: 10.1.1.21.3818.