



9th Applied Business and Engineering Conference

KLASIFIKASI CITRA MAMMOGRAM MENGGUNAKAN METODE K-MEANS CLUSTERING, GLCM, DAN *SUPPORT VECTOR MACHINE*(SVM)

Jihan Tiara Amanda¹⁾, Wahyuni Khabzli²⁾

¹Program Studi Teknil Elektronika Telekomunikasi, Politeknik Caltex Riau, Jl. Umban Sari No. 1, Pekanbaru, 28265

² Program Studi Teknil Elektronika Telekomunikasi, Politeknik Caltex Riau, Jl. Umban Sari No. 1, Pekanbaru, 28265

E-mail: jihan17tet@mahasiswa.pcr.ac.id

Abstract

Breast cancer is one of the non-communicable diseases that tend to continue to increase every year. The disease occurs almost entirely in women, but can also occur in men. The best way to identify the presence of breast cancer in addition to ultrasound examination can also be by interpreting a mammogram image using low doses of X-rays that can show abnormalities or abnormalities in the breast in a very small form. The cancer detection system that will be built is a detection system in the breast using mammogram imagery that will pass through the pre-processing stage, image segmentation stage, post-processing stage, feature extraction stage, and classification stage. Methods in the stages of breast cancer detection system are K-Means Clustering Method in segmentation process, GLCM in feature extraction and SVM in classification process. This system will detect and classify normal or abnormal breast cancer based on the characteristics that have been extracted, namely contrast, correlation, energy, and homogeneity by using the process of training (training) and testing (test). The accuracy achieved on this system is 85% of the 20 test images attempted.

Keywords: *Mammogram image, K-Means, GLCM, SVM*

Abstrak

Kanker payudara termasuk salah satu penyakit tidak menular yang cenderung terus meningkat setiap tahunnya. Penyakit ini terjadi hampir seluruhnya pada wanita, tetapi dapat juga terjadi pada pria. Cara terbaik untuk mengidentifikasi keberadaan kanker payudara selain pemeriksaan menggunakan Ultrasonografi (USG) dapat juga dengan menafsirkan gambar mammogram yang menggunakan sinar-X dengan dosis rendah yang dapat memperlihatkan keabnormalan atau kelainan pada payudara dalam bentuk yang sangat kecil. Sistem pendeteksi kanker yang akan dibangun merupakan sistem deteksi pada bagian payudara dengan menggunakan citra mammogram yang akan melewati tahap *pre-processing*, tahap segmentasi citra, tahap *post-processing*, tahap ekstraksi ciri, dan tahap



9th Applied Business and Engineering Conference

klasifikasi. Metode pada tahap-tahap sistem pendeteksian kanker payudara adalah Metode *K-Means Clustering* dalam proses segmentasi, GLCM dalam ekstraksi ciri dan SVM dalam proses klasifikasi. Sistem ini akan mendeteksi dan mengklasifikasikan penyakit kanker payudara normal atau abnormal berdasarkan ciri-ciri yang telah diekstraksi yaitu *contrast*, *correlation*, *energy*, dan *homogeneity* dengan menggunakan proses *training*(latih) dan *testing*(uji). Akurasi yang dicapai pada sistem ini adalah 85% dari 20 citra uji yang dicobakan.

Kata Kunci: *Citra Mammogram, GLCM, K-Means, SVM*

PENDAHULUAN

Penyakit kanker merupakan salah satu terbesar penyebab utama kematian pada manusia. Kanker termasuk golongan penyakit yang timbul karena pertumbuhan sel yang tidak terkontrol yang berpotensi untuk menyerang atau menyebar ke bagian lain dari tubuh. Terdapat lebih dari 100 jenis kanker yang berbeda yang masing-masing dibagi berdasarkan sel awal yang terbentuk. Faktor-faktor yang dapat menimbulkan kanker antara lain adalah mengkonsumsi tembakau, obesitas, kurangnya aktifitas fisik dan mengkonsumsi alkohol.

Kanker payudara termasuk salah satu penyakit kanker tidak menular yang cenderung terus meningkat setiap tahunnya. Penyakit ini terjadi hampir seluruhnya pada wanita, tetapi dapat juga terjadi pada pria. Kanker payudara menempati urutan kedua pada perempuan setelah kanker rahim. Kanker payudara merupakan penyebab kematian utama diantara wanita yang berusia antara 35-50 tahun.(Junita, no date)

Setiap tahun American Cancer Society memperkirakan jumlah kasus kanker baru dan kematian yang akan terjadi di Amerika Serikat pada tahun berjalan dan mengkompilasi data terbaru pada kejadian kanker, kematian, dan kelangsungan hidup. Berdasarkan *Estimated New Female Breast Cancer Cases and Deaths by Age, US, 2017* angka kasus *Carcinoma In Situ* (kanker payudara tahap awal) untuk segala umur mencapai 63.410 kasus, untuk *kasus Carcinoma Invasive* (kanker payudara yang sudah pecah menyebar ke organ–organ lain) mencapai 252.710 kasus dan untuk angka kematian mencapai 40.610. Menurut data GLOBOCAN (IARC) tahun 2012 diketahui

1000

ISSN: 2339 – 2053

Pekanbaru, 25 Agustus 2021



9th Applied Business and Engineering Conference

bahwa kanker payudara merupakan penyakit kanker dengan presentase kasus baru (setelah dikontrol oleh umur) tertinggi, yaitu sebesar 43,3% dan pesentase kematian (setelah dikontrol oleh umur) akibat kanker payudara sebesar 12,9%.

Dengan adanya data tersebut, maka perlu sekali bagi perempuan untuk mengetahui cara mendeteksi kanker payudara agar dapat dicegah. Skrining (screening) merupakan suatu cara untuk mendeteksi kanker stadium awal dan sangat membantu untuk meningkatkan persentase kesembuhan pada penderita kanker. (Tianur, Nugroho, Muzni Sahar, et al., 2017). Pengobatan kanker payudara pada stadium awal secara umum memberi hasil yang lebih baik dibanding stadium lanjut, oleh karena itu perempuan sebaiknya memahami tentang kanker payudara sehingga bisa mendeteksi secara dini terhadap gejalagejala terjadinya kanker payudara. (Junita, no date)

Metode skrining terbagi atas 3(tiga) yaitu Mammografi, *Magnetic Resonance Imaging* (MRI) dan USG. Selain USG yang merupakan metode paling sering digunakan untuk mendiagnosis kanker payudara, cara terbaik untuk mengidentifikasi keberadaan kanker payudara dapat juga dengan menafsirkan gambar mammografi yang menggunakan sinar-X yang dapat memperlihatkan keabnormalan atau kelainan pada payudara dalam bentuk yang sangat kecil. Jika skrining mamografi dilakukan secara teratur, dapat memberikan keuntungan pada wanita yang berusia dibawah 50 tahun. Berdasarkan Sistem Informasi RS (SIRS), jumlah penderita kanker payudara di Indonesia (28.7%).

Pendeteksian kanker payudara ini dengan menggunakan Metode K-Means Clustering dalam proses segmentasi, GLCM dalam ekstraksi ciri dan SVM dalam proses klasifikasi. Dimana citra yang akan diinput melewati proses pra-pengolahan atau preprocessing yang bertujuan untuk meningkatkan kualitas citra yang dihasilkan agar dapat memberikan informasi yang lebih jelas sehingga dapat mempermudah dalam melakukan interpretasi atas suatu citra. Teknik *preprocessing* yang digunakan yaitu Region of Interest (ROI), ROI digunakan secara otomatis yang dapat memangkas atau



9th Applied Business and Engineering Conference

memotong sebagian citra Mammogram pada koordinat tertentu untuk mengesktrak area yang mengandung lesi kanker payudara pada citra hasil Mammogram yang dimasukkan. *Tools* yang akan digunakan untuk mengerjakan Proyek Akhir ini ialah Software Matlab. Hasil yang akan diperoleh pada Proyek Akhir ini ialah memastikan, apakah Citra Mammografi yang diinputkan normal atau tidak. Sesuai dengan database yang diperoleh dari *Department of Physics, Royal Marsden Hospital, Fulham Road, London*.

Adapun tujuan dari penelitian ini adalah mengklasifikasi citra hasil Mammografi pada area kanker/tumor payudara normal atau abnormal.

METODE PENELITIAN

Metode penelitian yang digunakan pada tugas akhir ini adalah:

a. Studi Literatur

Pada tahap ini dilakukan studi literatur berupa buku referensi, artikel, dan jurnal-jurnal yang dapat menunjang Proyek Akhir ini.

b. Konsultasi dan diskusi

Tahap ini dilakukan dengan dosen pembimbing yang kompeten terhadap Pengolahan Citra menggunakan Matlab dan Weka.

c. Pembuatan Sistem

Melakukan pembuatan sistem yang digunakan untuk melakukan pendeteksian citra. Hasil yang didapat kemudian dianalisa dan di tuangkan dalam bentuk laporan.

d. Penulisan Laporan

Dalam penulisan laporan ini mengacu padapedoman penulisan ilmiah, dalam hal ini penulisan Proyek Akhir yang bentuk bakunya telah diatur oleh pihak Politeknik Caltex Riau.

HASIL DAN PEMBAHASAN

Adapun hasil proses terakhir dalam mendeteksi citra mammogram yaitu klasifikasi. Klasifikasi pada sistem ini dilakukan dengan menggunakan metode *Support Vector*

1002

ISSN: 2339 – 2053

Pekanbaru, 25 Agustus 2021



9th Applied Business and Engineering Conference

Machine (SVM) yang mana SVM ini merupakan salah satu metode dalam *supervised learning* yang sering digunakan untuk klasifikasi. SVM memiliki konsep klasifikasi yang lebih baik dan lebih jelas secara matematis dibanding metode-metode klasifikasi lainnya.

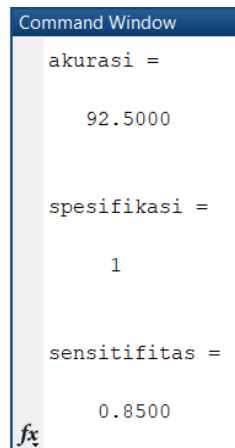
Pada proses klasifikasi ini, citra Mammogram nodul kanker payudara dikelompokkan berdasarkan Abnormal dan Normal. Proses pengolahan citra Mammogram dari tahap awal hingga tahap akhir sangat berpengaruh terhadap besarnya nilai akurasi klasifikasi yang diperoleh. Hal tersebut dikarenakan, semakin baik hasil segmentasi yang didapatkan dan juga semakin baik hasil proses *post-processing* yang dihasilkan maka hasil ekstraksi ciri-ciri fitur yang digunakan dari proses sebelumnya akan sangat berpengaruh untuk mendapatkan nilai akurasi klasifikasi yang tinggi. Semakin besar akurasi pada data pelatihan, maka semakin besar pula akurasi pada data uji.

Hasil ekstraksi ciri yang telah diperoleh dari tahap sebelumnya akan dilanjutkan ke proses klasifikasi. Pada semua program, proses klasifikasi SVM dilakukan hanya pada saat pelatihan/*training* data latih. Hal ini dilakukan agar citra yang ingin diuji sudah memiliki acuan dari hasil pelatihan data latih. Berikut merupakan *syntax* SVM pada M-file yang digunakan untuk menjalankan dan menampilkan hasil klasifikasi menggunakan metode SVM pada data latih atau saat *training* data:

```
Mdl = fitsvm(data_latih,target_latih,'Standardize',true,...  
            'KernelFunction','rbf','KernelScale','auto');
```

Pada *syntax* diatas, terdapat fungsi **fitsvm** yang mana fungsi dalam sistem ini digunakan untuk klasifikasi SVM dengan model klasifikasinya 2 kelas (*2 classes*). Jika model kelasnya lebih dari 2 kelas, fungsi yang dapat digunakan yaitu fungsi **multisvm**. Saat program ini dijalankan, hasil ekstraksi data abnormal dan normal sebelumnya akan di-*load*, lalu fungsi SVM dilakukan dan model klasifikasi ini akan disimpan untuk

menjadi acuan pada GUI. Pada Gambar 1 dibawah ini merupakan nilai akurasi klasifikasi pada **data latih** yaitu 92.5% untuk spesifikasi dan sensitifitasnya ialah 100% dan 85%.



```
Command Window
akurasi =
    92.5000

spesifikasi =
    1

sensitifitas =
    0.8500
fx
```

Gambar 1. Nilai Akurasi, Spesifikasi, dan Sensitifitas pada Data Latih

Pada GUI, *syntax* yang digunakan untuk klasifikasi hanya memanggil atau *load* file data latih yang sebelumnya sudah disimpan untuk dijadikan acuan. Berikut adalah *syntax* klasifikasi pada GUI:

```
set(handles.edit2,'String','')
ciri = handles.ciri;
load Mdl
kelas_keluaran = predict(Mdl,ciri);
set(handles.edit2,'String',kelas_keluaran)
```

Load Mdl pada *syntax* diatas merupakan *syntax* untuk memanggil M-file yang Bernama Mdl. Untuk hasil pengujian, dari 20 citra **data uji** yang dicobakan pada GUI, hasil akurasi, spesifikasi, dan sensitifitasnya adalah dapat dilihat pada Gambar 85%, 90%, dan 80%.

```
Command Window
akurasi =
    85
spesifikasi =
    0.9000
sensitifitas =
    0.8000
```

Gambar 2. Nilai Akurasi, Spesifikasi, dan Sensitifitas pada Data Uji

Untuk memastikan hasil klasifikasi benar atau tidak, hasil klasifikasi citra mammogram ini dapat dibandingkan dengan keterangan yang sudah ada pada *database* dari *Department of Physics, Royal Marsden Hospital, Fulham Road, London*.

KESIMPULAN

Dari data yang didapat dan analisis yang telah dilakukan maka dapat ditarik kesimpulan yaitu :

1. Citra Mammogram dapat dideteksi menggunakan metode K-Means Clustering untuk segmentasi, GLCM untuk ekstraksi ciri, dan SVM untuk klasifikasi.
2. Sistem ini telah dirancang untuk memberikan informasi mengenai penyakit payudara yang teridentifikasi sebagai normal atau abnormal.
3. Hasil klasifikasi citra mammogram payudara menggunakan SVM menunjukkan hasil yang baik dan memperoleh nilai akurasi mencapai 85%.

Dari penelitian yang telah dilakukan, Adapun saran untuk pengembangan tugas akhir pada penelitian selanjutnya, yaitu:



9th Applied Business and Engineering Conference

1. Database citra mammogram yang digunakan pada sistem ini adalah citra mammogram lama. Untuk citra mammogram terbaru dapat menggunakan sistem terbaru pula.
2. Untuk pengembangan penelitian selanjutnya sistem dapat dibuat dalam bentuk aplikasi sehingga dapat diakses dilaptop yang tidak memiliki software Matlab.

DAFTAR PUSTAKA

- Tianur, Nugroho, H. A., Sahar, M., Indrastuti, R., & Choridah, L. (2017). Classification of Breast Ultrasound Images based on Posterior Feature. *International Conference on Information Technology Systems and Innovation (ICITSI), October, 5–6.*
- Amalia, A. E., Airlangga, G., & Thohari, A. N. A. (2018). Breast Cancer Image Segmentation Using K-Means Clustering Based on GPU Cuda Parallel Computing. *JURNAL INFOTEL*. <https://doi.org/10.20895/infotel.v10i1.344>
- Dhanachandra, N., Mangle, K., & Chanu, Y. J. (2015). Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2015.06.090>