EKSTRAKSI CIRI BERBASIS WAVELET DAN GLCM UNTUK DETEKSI DINI KANKER PAYUDARA PADA CITRA MAMMOGRAM

Hanifah Rahmi Fajrin[1*], Hanung Adi Nugroho[2], Indah Soesanti[3]
Jurusan Teknik Elektro dan Teknologi informasi Universitas Gadjah Mada
Yogyakarta, Indonesia
Email : hanifah.fajrin.sie13@mail.ugm.ac.id[*], adinugroho@ugm.ac.id, indsanti@gmail.com

Abstrak

Kanker payudara merupakan pembunuh nomor dua di dunia setelah kanker mulut rahim pada wanita. Dengan adanya deteksi dini kanker payudara kesempatan untuk bertahan hidup bagi penderita dapat ditingkatkan. Pada penelitian ini dilakukan pengolahan citra yang dapat melakukan pendeteksian dini terhadap kanker payudara. Terlebih dahulu dilakukan pra pengolahan pada citra dengan median filter dan connected component labeling (CCL) yang bertujuan untuk meningkatkan kualitas dan menghilangkan derau pada mammogram. Dengan mengekstrak ciri energi dari wavelet dekomposisi "haar" level 3, entropi, dan juga 5 ciri GLCM : IDM, ASM, korelasi, entropi, kontras. kemudian dilakukan klasifikasi berbasis statistik yaitu dengan regresi logistik untuk mendeteksi apakah citra mammogram termasuk normal atau abnormal. Penelitian dilakukan pada 108 data, yaitu 78 data abnormal dan 30 data normal, untuk pengujian dilakukan dengan algoritma k-fold validation. Pada fold-11 didapatkan nilai akurasi 81,45%, sensitivitas 82% dan spesifisitas 77,78%.

Kata Kunci : GLCM, regresi logistik., transformasi wavelet

I. PENDAHULUAN

Kanker adalah suatu penyakit dimana terjadi pertumbuhan berlebihan atau perkembangan tidak terkontrol dari sel-sel jaringan pada bagian tubuh tertentu. Kanker payudara merupakan jenis kanker yang sering ditemukan pada wanita. Menurut WHO pada tahun 2008 ada sekitar 1,38 juta kasus terbaru dan 458.000 wanita meninggal tiap tahunnya diakibatkan oleh kanker payudara, lebih dari setengahnya, yaitu sekitar 269.000 terdapat di negara berkembang dengan angka pendapatan perkapita yang rendah (WHO, 2008). Sedangkan di Indonesia menurut profil kesehatan Departemen Kesehatan Republik Indonesia Tahun 2012 kanker tertinggi yang diderita wanita Indonesia adalah kanker payudara dengan angka kejadian 2.2 % dari 1000 perempuan. Jika hal ini tidak bisa terkendali, maka diperkirakan pada tahun 2030 akan ada 26 juta orang menderita kanker payudara dan 17 juta meninggal dunia(DEPKES, 2012). Untuk itu diperlukan adanya deteksi dini kanker payudara melalui mamografi untuk dapat meningkatkan kesempatan bertahan hidup. Hal tersebut terbukti di Amerika dan Inggris dengan deteksi dini dapat menyelamatkan 12 sampai 37 jiwa perhari. Dengan perspektif, di Amerika dari 527 kasus kanker payudara dengan tingkat kematian 110 perhari. Di Inggris dari 125 kasus kanker payudara dengan tingkat kematian 35 perhari, jika dilakukan deteksi dini dapat menyelamatkan 12 jiwa per hari (WHO, 2008).

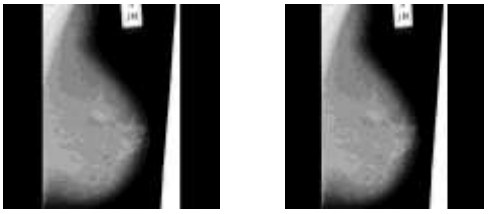
Mammogram merupakan proses skrining dalam bidang kedokteran yang digunakan untuk menemukan tumor payudara. Dan merupakan skrining yang paling umum digunakan untuk kanker payudara. *Mammogram* juga telah terbukti efektif dalam melihat ada atau tidaknya tumor pada payudara (M. Vidhya, dkk, 2011). *Computer-Aided Diagnosis* adalah sebuah sistem yang mampu mendiagnosis atau dalam arti sebenar nya dapat membedakan adanya penyakit atau tidak, dan mengurangi tingkat kesalahan dari pembacaan *false positive* dan *false negative*, serta meningkatkan peluang untuk mendeteksi adanya keadaan abnormal lebih dini. *False positive* merupakan suatu kondisi sakit yang dideteksi sebagai kondisi sehat sedangkan *false negative* suatu kondisi sehat yang dideteksi sebagai sakit. Kondisi yang diharapkan yaitu citra terbaca *sebagai True positive* yaitu kondisi sakit yang positif dideteksi sebagai sakit dan *True Negative*, yaitu kondisi tidak sakit (sehat) dan positif dideteksi sebagai sehat.

II. METODOLOGI PENELITIAN

2.1 Pra pengolahan

Pada tahap pra pengolahan, akan dilakukan hal-hal yang bertujuan untuk mempermudah proses pengolahan citra untuk tahap selanjutnya. Pada penelitian ini, citra di tapis dengan tapis median terlebih

dahulu. Tapis *median* digunakan untuk mengurangi derau *"salt and pepper"* dan untuk *smoothing* pada citra *mammogram. Smoothing* merupakan suatu teknik yang bertujuan untuk memperhalus gambar, serta mengurangi *derau*. Berikut hasil proses pra pengolahan :



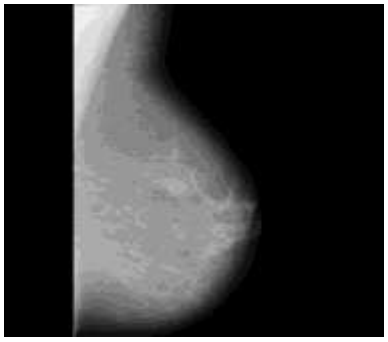
(a) (b)

Gambar 1. Citra hasil pra pengolahan (a). citra asli, (b). citra hasil median filter

Selanjutnya dengan algoritma *connected component labeling* dilakukan proses untuk menghilangkan label atau artefak yang terdapat pada citra. *connected component labeling* ini merupakan suatu algoritma pelabelan pada citra biner. Pelabelan terhadap citra ini merupakan tindakan memberikan label yang berbeda pada citra, dengan ketentuan (T. Roorkee,2010):

$$B(y,x) = \begin{cases} 0 & piksel\ latar\ belakang \\ 1 & piksel\ latar\ depan \\ 2,3,\ldots & label\ objek \end{cases} \quad \ldots(2.1)$$

Setelah didapatkan objek dengan label yang berbeda maka dilakukan proses untuk menghilangkan objek tersebut.



Gambar 2. Citra hasil CCL (*Connected Component Labeling)*

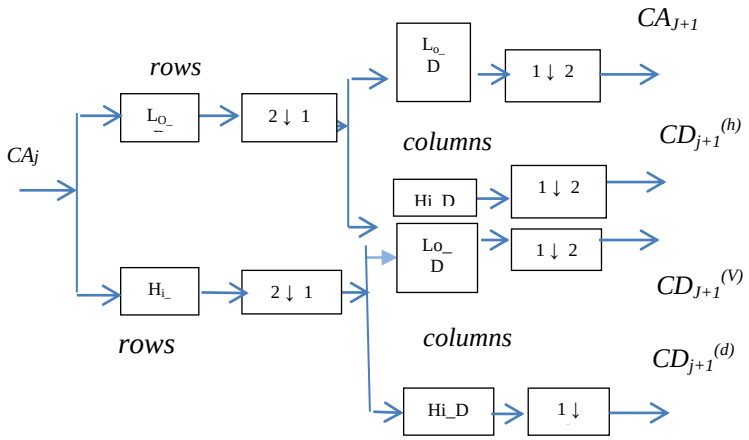
2.2 Ekstraksi Ciri

Ekstraksi ciri yaitu proses untuk mendapatkan ciri yang mewakili citra itu sendiri. Ciri-ciri tersebut akan digunakan untuk mengenali suatu objek sesuai dengan kategorinya. Disini metode ekstraksi ciri dengan memakai GLCM, transformasi wavelet dan entropy. Ciri-ciri yang telah diekstrak menjadi masukan pada tahap klasifikasi.

2.2.1 Dekomposisi Wavelet

Implementasi *transformasi wavelet diskrit* dapat dilakukan dengan cara melewatkan sinyal frekuensi tinggi atau *highpass filter* dan frekuensi rendah atau *lowpass filter* (S.Amutha, 2005)

ISBN 978-602-99334-4-4



Gambar 3. transformasi wavelet diskrit dua dimensi dengan level dekomposisi satu

Dimana :

2 ↓ 1	Merupakan downsample kolom
1 ↓ 2	Merupakan downsample baris

Seperti yang terlihat pada gambar , jika suatu citra dilakukan proses *transformasi wavelet diskrit* dua dimensi dengan *level dekomposisi* satu, maka akan menghasilkan empat buah *subband,* yaitu (Irtawaty, Andi Sri, 2014) :

1. *Koefisien Approksimasi* (CA j+1) atau disebut juga *subband* LL
2. *Koefisien Detil Horisontal* (CD(h) j+1) atau disebut juga *subband* HL
3. *Koefisien Detil Vertikal* (CD(v) j+1) atau disebut juga *subband* LH
4. *Koefisien Detil Diagonal* (CD(d) j+1) atau disebut juga *subband* HH

Ciri yang diambil adalah ciri energi pada *wavelet*. Energi merepresentasikan keseragaman tekstur dari citra. Nilai energi sendiri diambil dari 4 (empat) nilai-nilai *koefisien aproksimasi* (ca), *koefisien detail* arah *horizontal* (ch), koefisien detail arah *vertical* (cv), dan *koefisien detail* arah *diagonal* (cd) yang nilainya tergantung pada nilai *wavele*t-nya.Energi dibagi dalam emapat ciri, yaitu (Andi Sri, 2014) :

1. Energi yang berhubungan dengan nilai pendekatan *(aproksimasi)/Ea,* Ea dihitung berdasarkan persentase jumlahan kuadrat dari nilai *koefisien aproksimasi* ca dibagi dengan jumlahan seluruh koefisien c.

$$E_a = \frac{\sum c_a^2}{\sum c^2} x\ 100\% \text{.....(2.2)}$$

2. Energi yang berhubungan dengan nilai detail pada arah horizontal/Eh, Eh dihitung berdasarkan persentase jumlahan kuadrat dari nilai koefisien detail pada arah horizontal ch dibagi dengan jumlahan seluruh koefisien c.

$$E_h = \frac{\sum c_h^2}{\sum c^2} x\ 100\% \text{.....(2.3)}$$

3. .Energi yang nilai berhubungan dengan nilai detail pada arah vertikal/Ev, Ev dihitung berdasarkan persentase jumlahan kuadrat dari nilai koefisien detail pada arah horizontal chorizontal cv dibagi dengan jumlahan seluruh koefisien c.

$$E_v = \frac{\sum c_v^2}{\sum c^2} x\ 100\% \text{.....(2.4)}$$

4. Energi yang berhubungan dengan nilai detail pada arah diagonal/Ed, Ed dihiitung berdasarkan persentase jumlahan kuadrat dari nilai koefisien detail pada arah diagonal cd dibagi dengan jumlahan seluruh koefisien c.

$$E_d = \frac{\sum c_d^2}{\sum c^2} x\ 100\% \text{.....(2.5)}$$

2.2.2 *Entropi*

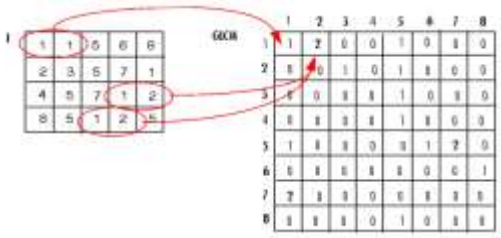
Entropi merepresentasikan sifat keacakan dari suatu citra. Sifat keacakan sangat terkait erat dengan informasi. Analisis dapat dilakukan dengan berbasis pada informasi yang diperoleh berdasar sifat keacakan ini. Shannon mendefinisikan entropi secara matematis pada persamaan dasar berikut :

$$H = -\sum_{i=0}^{N} p_i \,.\log(p_i) \ldots..(2.6)$$

dengan H adalah entropy dan pi adalah probabilitas kejadian i. Untuk data digital digunakan log dengan bilangan dasar 2. Entropy juga mendefinisikan rerata informasi yang diperoleh.

2.2.3 GLCM

GLCM merupakan ekstraksi ciri statistik orde kedua dengan memakai matriks kookurensi, yaitu suatu matriks antara yang merepresentasikan hubungan ketetanggaan antar piksel dalam citra pada berbagai arah orientasi dan jarak spasial. Salah satu teknik untuk memperoleh ciri statistik orde dua adalah dengan menghitung probabilitas hubungan ketetanggaan antara dua piksel pada jarak dan orientasi sudut tertentu. Pendekatan ini bekerja dalam dua tahapan, yaitu: pembentukan sebuah matriks kookurensi dari data citra, dilanjutkan dengan penghitungan ciri sebagai fungsi dari matriks antara tersebut. Kookurensi berarti kejadian bersama, yaitu jumlah kejadian satu level nilai piksel bertetangga dengan satu level nilai piksel lain dalam jarak (*d*) dan orientasi sudut (θ) tertentu. Jarak dinyatakan dalam piksel dan orientasi dinyatakan dalam derajat. Orientasi dibentuk dalam empat arah sudut dengan interval sudut 45°, yaitu 0°, 45°, 90°, dan 135°. Sedangkan jarak antar piksel biasanya ditetapkan sebesar 1 piksel. Matriks kookurensi merupakan matriks bujursangkar dengan jumlah elemen sebanyak kuadrat jumlah level intensitas piksel pada citra. Setiap titik (*p*,*q*) pada matriks kookurensi berorientasi θ berisi peluang kejadian piksel bernilai *p* bertetangga dengan piksel bernilai *q* pada jarak *d* serta orientasi θ dan (180−θ)(Zulfah tri, 2012). Fitur GLCM yang dipakai disini adalah *Angular Second Moment, Contrast, Correlation, Inverse Different Moment, Entropy* .



Gambar 4. Ilustrasi GLCM

2.3 Klasifikasi Payudara Normal Dan Abnormal

Regresi logistik adalah salah satu model untuk menduga hubungan antara peubah respon kategori dengan satu atau lebih peubah prediktor yang kontinyu ataupun kategori. Peubah respon yang terdiri dari dua kategori yaitu "ya (sukses)" dan "tidak (gagal)", dan dinotasikan 1="sukses" dan 0="gagal", maka akan mengikuti sebaran Bernoulli. Model regresi logistic dinyatakan (fitrianti, dkk, 2014) :

$$\pi(X_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{12} + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{12} + \beta_p x_{ip})} \ldots\ldots(2.7)$$

Proses pendugaan parameter dari regresi logistik menggunakan metode MLE. Metode MLE memberikan nilai duga bagi β dengan cara memaksimumkan fungsi likelihood dan mensyaratkan bahwa data mengikuti sebaran Bernoulli. Fungsi likelihood untuk model regresi logistik dikotomus adalah (fitrianti, dkk, 2014):

$$\xi(\beta) = \prod_{i=1}^{n} f\,(\beta, y) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{t-y_i} \ldots.(2.8)$$

Agar nilai fungsi mencapai maksimum maka turunan parsial pertama terhadap β_j disamakan dengan nol. Persamaan hasil turunan masih nonlinier, maka dibutuhkan metode iterasi Newton-Raphson [12]. Pengujian signifikansi parameter model regresi logistik dilakukan secara simultan dan secara parsial.Pengujian secara simultan dilandaskan pada hipotesis :

ISBN 978-602-99334-4-4

H_0: $\beta_i = \beta_2 = \cdots = \beta_j = 0$ (tidak ada pengaruh antara peubah prediktor terhadap peubah respon)
H_1: paling sedikit ada satu $\beta_j \neq 0$(ada pengaruh antara peubah prediktor terhadap peubah respon) dengan statistik uji G adalah:

$$-2 \ln \left[\frac{L_0}{L_1}\right] \sim X^2(p) \ldots\ldots (2.9)$$

Statistik uji-G mengikuti sebaran X^2 dengan derajat bebas sama dengan banyaknya parameter β_j, di mana H_0 akan ditolak jika nilai statistik uji G $\geq X^2_{(p,0.05)}$ dengan tingkat kepercayaan (1-α)100. Sedangkan pengujian secara parsial dilandaskan pada hipotesis:
H_0: $\beta_j = 0$ (tidak ada pengaruh antara masing-masing peubah prediktor terhadap peubah respon)
H_1: $\beta_j \neq 0$ (ada pengaruh antara masing-masing peubah prediktor terhadap peubah respon)
Rumus statistik uji Wald adalah :

$$\left[\frac{\beta_j}{Se\,(\beta_j)}\right] \sim Z \quad ; j = 0, 1, 2, \ldots\ldots, p \ldots\ldots (2.10)$$

Hipotesis nol ditolak jika $|W| > Z_{\alpha/2}$artinya peubah prediktor berpengaruh nyata terhadap peubah respon (Hosmer dan Stanley, 2000). Hosmer dan Stanley (2000) menyatakan bahwa peubah respon dengan dua kategori (biner) dengan ketentuan jika $\pi(x) \geq 0.5$maka hasil prediksi adalah 1, jika $\pi(x) < 0.5$ maka hasil prediksi adalah 0. Klasifikasi menggunakan model peluang dengan persamaan sebagai berikut (rahman farisi, 2012):

$$logit\ \pi(x_i) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \ldots (2.11)$$

Pada penelitian ini variable respon yang digunakan adalah kategori apakah citra mammogram termasuk kedalam normal atau abnormal berdasarkan nilai ciri yang didapatkan setelah proses ekstraksi ciri. Sedangkan variable prediktornya berupa ciri GLCM, entropi dan energy wavelet yaitu ada 10 ciri.

III. HASIL DAN PEMBAHASAN

Setelah melakukan ekstraksi ciri dengan fitur energy wavelet, entropi dan GLCM selanjutnya dilakukan proses klasifikasi dengan menggunakan machine learning weka dengan menggunakan metode regresi logistic, yang merupakan analisis klasifikasi berbasis statistic. Dari confussion matriks pada weka Pada 108 data dari mias dataset dan teknik pengujian dengan croos validation. Untuk mendapatkan nilai akurasi, sensitivity dan specificity dihitung dengan rumus (Tri M, Zulfa, 2012) :

$$\%\ akurasi = \frac{TP+TN}{TP+FN+TN+FN}\ x\ 100 \ldots.(3.1)$$

$$\%\ sensitivitas = \frac{TP}{TP+FN}\ x\ 100 \ldots..(3.2)$$

$$\%\ spesitifitas = \frac{TN}{TN+FP} x\ 100 \ldots..(3.3)$$

Dari confussion matriks pada weka nilai akurasi, sensitivity dan specificity dapat dihitung sebagai berikut :

$$\%\ akurasi = \frac{74+14}{74+14+4+16}\ x\ 100\% = 81.4815\%$$

$$\%\ sensitifity = \frac{74}{74+16}\ x\ 100\% = 82\%$$

$$\%\ specificity = \frac{14}{14+4}\ x\ 100\% = 77{,}78\%$$

Dari hasil pengujian didapatkan nilai akurasi sebesar 81.4815%, sensititivitas 82% dan spesifisitas 77,78% pada fold ke 11.

Gambar 5. Hasil klasifikasi dengan metode regresi logistic dengan tools weka

IV. KESIMPULAN

1. Ciri energi hasil dekomposisi wavelet, entropi dan GLCM dapat dijadikan sebagai ciri yang akan diektrak sebagai inputan untuk proses klasifikasi karena merepresentasikan keberagaman tekstur dan sifat keacakan dari citra.
2. System logistik regresi merupakan analisis klasifikasi berbasis statistik yang dapat dijadikan sebagai system klasifikasi pada klasifikasi antara payudara normal.

DAFTAR PUSTAKA :

M. Vidhya, N. Sangeetha, and M. N. Vimalkumar, "Early Stage Detection of Cancer in Mammogram Using Statistical Feature Extraction," pp. 401–404, 2011.

S. A. Butler, R. J. Gabbay, D. A. Kass, D. E. Siedler, K. F. O. Shaughnessy, and R. A. Castellino, "Computer-Aided Detection in Diagnostic Mammography :," no. November, pp. 1511–1515, 2004.

T. Roorkee, "Contrast Enhancement of Mammographic Images Using Wavelet Transform Vijaya Kumar Gunturu Ambalika Sharma," pp. 323–327, 2010.

Rachman, Farizi dan Purnami, Santi Wulan. Perbandingan klasifikasi tingkat keganasan breast cancer dengan menggunakan regresi logistic ordinal dan support vector machine. Jurnal sains dan seni ITS vol 1. 2012

Fitrianty, Delbra Andhini, S. Wardhani, Ni Wayan dan A. Soebono Loekito. Ketepatan Klasifikasi dengan Analisis Regresi Logistic dan Multivariate Adaptive Regression Splines (MARS) pada Data dengan Peubah Respon Biner. Jurusan matematika universitas Brawijaya. 2014

www.who.int/cancer/event/breast_cancer_month/en/

www.depkes.go.id/index.php/press-realise/

M. Transformasi and P. Wavelet, "Makalah seminar tugas akhir analisis tekstur menggunakan metode transformasi paket," pp. 1–8, 1982.

M. T. Akhir, "Identifikasi keberadaan kanker pada citra mammografi menggunakan metode wavelet haar," pp. 1–7.

S. Amutha and R. S. M, "Bi-Orthogonal Wavelet."

T. Discrete and W. Transform, "Topic 7 : Discrete Wavelet Transform."

Irtawaty, Andi Sri " ekstraksi ciri berbasis wavelet untuk perhitungan kuantitas bakteri actobacillus bulgarius dan streptococcus thermophiles pada yogurt'.universitas gadjahmada. 2014

Tri M, Zulfa.2012. Klasifikasi Kanker Usus berdasarkan Citra Mikroskopik Patologi Menggunakan *Contourlet Transform* dan *Support Vector Machine* (SVM). Skripsi. Universitas Telkom

ISBN 978-602-99334-4-4