

Analisis Regresi dan Analisis Diskriminan untuk Mengukur Tingkat Akurasi Feature Citra Termogram

Okky Dwi Nurhayati

Jurusan Sistem Komputer, Fakultas Teknik Universitas Diponegoro
Jl. Prof. H. Soedarto, Tembalang, Semarang
okky_0210@yahoo.co.id

Intisari — Analisis diskriminan adalah analisis multivariat yang diterapkan untuk memodelkan hubungan antara satu variabel respon yang bersifat kategori dengan satu atau lebih variabel prediktor yang bersifat kuantitatif. Sedangkan analisis regresi bertujuan untuk membentuk sebuah fungsi yang dapat menjelaskan hubungan dua variabel, yaitu variabel penjelas/prediktor (x) dan variabel respon (y). Banyak aplikasi pada bidang kedokteran atau industri yang berhubungan dengan data mining salah satunya untuk pengenalan pola pada citra termogram.

Tujuan dari penelitian ini adalah membandingkan teknik analisis diskriminan atau linear discriminant analysis (LDA) dan analisis regresi pada tingkat akurasi pengenalan pola citra termogram. Penelitian ini menggunakan sampel citra digital termogram payudara yang diambil dari kamera Fluke Ti20. Jumlah sampel yang digunakan adalah 60 citra termogram yang di bagi masing-masing ke dalam tiga kelas yaitu kelas normal, kelas kanker payudara dini, dan kelas payudara lanjut.

Dari penelitian yang telah dilakukan dapat dibuktikan bahwa analisis diskriminan dengan 2 feature (ciri), 3 ciri, dan 5 ciri pada citra termogram memberikan tingkat akurasi 81,7 %. Sedangkan analisis diskriminan dengan 4 ciri pada citra termogram memberikan tingkat akurasi yang paling tinggi yaitu 83,33 %.

Kata kunci — termogram, multivariat, kovarian, ciri, analisis diskriminan

Abstract — Discriminant analysis is a multivariate analysis applied to model the relationship between the response variable is the category with one or more predictor variables that are quantitative. While regression analysis aims to establish a function that can explain the relationship between two variables, namely the explanatory variables / predictors (x) and the response variable (y). Many applications in the medical field or industry related to one of data mining for pattern recognition in the thermogram image.

The aim of this study is to prove the technique of linear discriminant analysis (LDA) and regression analysis to distinguish the types of thermogram. This study used 60 samples of breast thermograms captured from camera Fluke Ti20. The samples used are images in the thermograms which each classify into three classes, namely breast normal thermogram, early breast cancer thermogram, and advanced breast cancer thermogram.

The result of research, discriminant analysis with two features, three features, and five features give 81.7% accuracy rate. While discriminant analysis with four features have the highest accuracy rate is 83.33%. Final results of the regression analysis is able to significantly separate the three types of normal, early, and advanced thermogram.

Keywords — thermogram, multivariate, covarian, feature, discriminat analysis, regression analysis

I. PENDAHULUAN

Pengolahan citra merupakan salah satu bidang yang berkembang cukup pesat saat ini. Banyak aplikasi pengolahan citra digunakan dalam kehidupan sehari-hari [3] Pengolahan

citra banyak diterapkan dalam bidang kedokteran untuk menganalisis, mendiagnosis, maupun mendeteksi citra medis berdasarkan hasil klasifikasi citra [4][5]. Berbagai aplikasi pengolahan citra

sangat membantu bagi kepentingan manusia. Salah satu aplikasinya adalah peningkatan kualitas citra medis sebagai pra-proses agar lebih mudah untuk dianalisis atau diinterpretasikan oleh visual manusia (tenaga kesehatan) [4][5].

Pengolahan citra berhubungan dengan visi komputer yang melibatkan operasi-operasi pengolahan citra mulai dari penangkapan citra oleh sistem kamera (system akuisisi data), pengolahan citra menjadi bentuk yang lebih ringkas dan sederhana namun masih tetap mewakili citra, sampai yang terpenting yaitu analisis citra untuk menentukan jenis citra/pengenalan pola [1][2].

Mengingat demikian pentingnya tahapan pe-nge-nalan pola dalam sistem visi komputer, maka pemilihan metode yang tepat pada masing-masing tahap perlu dilakukan terhadap sampel citra termogram. Beberapa algoritma pengenalan pola telah banyak dikembangkan. Pengenalan pola berhubungan dengan klasifikasi objek ke dalam suatu kelas tertentu [3]. Tiga metode klasifikasi yang sering di gunakan adalah klasifikasi melalui analisis Regresi, Analisis Diskriminan, dan Jaringan Syaraf Tiruan. Sedangkan metode ekstraksi citra yang digunakan sebagai input dari metode klasifikasi adalah metode statistik, metode struktural, maupun metode hibrid [3].

Pada penelitian ini menawarkan sebuah metode statistik yang digunakan untuk mengekstrak ciri objek pada analisis citra bertujuan untuk mengukur tingkat akurasi pengenalan citra termogram dalam suatu sistem visi komputer berdasarkan analisis diskriminan dan analisis regresi. Beberapa ciri statistik yang diekstrak dari sebuah citra diantaranya nilai rata-rata, standar deviasi, entropi, skewness, dan kurtosis [4].

II. LANDASAN TEORI

Analisis diskriminan adalah teknik statistik multivariate yang terkait dengan

pemisahan/klasifikasi sekelompok objek atau observasi ke dalam kelompok yang telah terlebih dahulu didefinisikan. Dalam tujuan pengenalan objek, metode ini mencoba menemukan suatu ‘discriminant’ yang nilainya secara numeris sedemikian sehingga mampu memisahkan objek yang karakteristiknya telah diketahui. Sedangkan dalam tujuan klasifikasi objek, metode ini akan mensortir objek (observasi) kedalam 2 atau lebih kelas [6].

Analisis diskriminan atau *linear discriminant analysis* (LDA) berbeda dengan analisis kluster dalam hal jenis masalah yang diselesaikan. Dalam LDA ada beberapa data yang sudah diketahui kelasnya atau labelnya. Data yang sudah diketahui labelnya dipergunakan untuk menemukan fungsi diskriminan [6]. Untuk data lain yang belum diketahui kelasnya/labelnya, dapat menggunakan fungsi diskriminan yang telah ditemukan.

Analisis ini didasarkan atas fungsi diskriminan yang mempunyai bentuk umum:

$$\hat{Y}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon_i \quad (1)$$

di mana Y adalah kelompok merupakan *dummy variables* dan Xi adalah variabel pembeda [6]. Pada dasarnya fungsi diskriminan merupakan fungsi regresi ganda dengan variabel dependen merupakan *dummy variables* yang mengambil nilai 1, 2, ..., k yang sesuai dengan pengelompokan awal setiap individu dan banyaknya kelompok (k).

Analisis diskriminan dapat dilakukan bila terdapat perbedaan yang nyata antar kelompok, sehingga pada tahap awal yang harus dilakukan adalah uji hipotesis nol bahwa tidak ada perbedaan kelompok di antara individu yang dirumuskan dengan persamaan-persamaan berikut:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (2)$$

$$H_1 : \mu_1 \neq \mu_2; j = 1, 2, \dots, k \quad (3)$$

$\mu_j = [\mu_{1j}, \mu_{2j}, \dots, \mu_{pj}]$ adalah vektor rata-rata hitung peubah pembeda pada populasi ke- j dan jumlah individu yang dianalisis $N = \sum N_i$ maka probabilitas pengelompokan awal (*prior probability*) adalah N_i/N , sedangkan k untuk menunjukkan kelompok.

Pada rumusan Linear Diskriminan Analisis (LDA) ini digunakan ide total *error classification* atau berarti metode ini akan berusaha untuk meminimalkan proporsi objek yang salah klasifikasi [6].

Ada dua kelompok ciri khas pada objek yaitu ciri global yang merupakan ciri khas keseluruhan objek dan ciri lokal yang merupakan ciri khas bagian tertentu dari objek[1][3]. Ciri yang akan diekstrak dari permukaan citra digital termogram payudara antara lain:

Momen (pusat objek) dari intensitas citra. Untuk menghitung posisi atau titik pusat suatu objek, digunakan persamaan-persamaan berikut:

$$\bar{x} = \frac{\sum_{i=1}^n \sum_{j=1}^m j \cdot f(i, j)}{A} \quad (4)$$

$$\bar{y} = \frac{\sum_{i=1}^n \sum_{j=1}^m i \cdot f(i, j)}{A} \quad (5)$$

dengan A adalah luas atau ukuran objek yang dihitung dengan persamaan:

$$A = \sum_{i=1}^n \sum_{j=1}^m j \cdot f(i, j) \quad (6)$$

Lima momen yang akan dihitung berdasarkan distribusi suhu citra antara lain: *mean*, *variance*, *skewness*, dan *kurtosis* [1][2]. Perhitungan masing-masing momen tersebut adalah:

$$a) \text{ mean } (\mu) = \frac{1}{N} \sum_{j=1}^N p_j \quad (7)$$

$$b) \text{ variance } \sigma^2 = \frac{1}{N-1} \sum_{j=1}^N (p_j - \mu)^2 \quad (8)$$

$$c) \text{ skewness } = \frac{1}{N} \sum_{j=1}^N \left(\frac{p_j - \mu}{\sigma} \right)^3 \quad (9)$$

$$d) \text{ kurtosis } = \frac{1}{N} \sum_{j=1}^N \left(\frac{p_j - \mu}{\sigma} \right)^4 \quad (10)$$

e) entropi yang diukur dari distribusi intensitas dengan persamaan berikut.

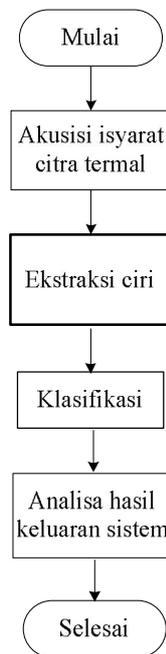
$$H(X, Y) = - \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} p_{ij} \log p_{ij} \quad (11)$$

III. MATERI DAN METODE PENELITIAN

Materi yang digunakan dalam penelitian ini adalah 60 sampel citra termogram digital yang di telah diakusisi menggunakan kamera thermal Fluke Ti20. Citra digital di download dari dalam kamera kemudian diatur ukuran citra menjadi 256 pikselx 256 piksel dengan format penyimpanan citra berekstensi .bmp. Dari jumlah sampel yang digunakan dalam penelitian, dilakukan pembagian sampel ke dalam 3 kelas berdasarkan jenis termogram, yaitu 20 sampel termogram normal, 20 sampel termogram kanker dini, dan 20 sampel termogram kanker lanjut.

Citra termogram digital di *download* dari kamera Fluke Ti 20 dengan format penyimpanan citra berekstensi .bmp untuk selanjutnya digunakan sebagai *input* pada program simulasi. Tahap selanjutnya adalah ekstraksi ciri statistik citra termogram dengan 5 ciri statistik yang diperoleh untuk digunakan sebagai *input* pada metode klasifikasi dengan menggunakan analisis diskriminan.

Penelitian ini dilakukan dengan langkah-langkah yang ditunjukkan pada Gambar 1 berikut.



Gbr. 1 Langkah-langkah penelitian

IV. HASIL DAN PEMBAHASAN

Setelah proses ekstraksi ciri dilakukan terhadap masing-masing citra, kemudian dilakukan pelabelan pada masing-masing citra berdasarkan label yang sudah diketahui. Kelas 1 untuk menunjukkan jenis citra termogram normal, kelas 2 untuk mewakili jenis termogram kanker dini, dan kelas 3 untuk mewakili jenis termogram kanker lanjut.

Hasil yang diperoleh dengan metode diskriminan linier dan kesalahan klasifikasi pada data *training* dan data *testing* citra termogram ditunjukkan pada Tabel 1 berikut.

Tabel 1 Tingkat akurasi citra termogram

Jumlah sampel/kelas	% akurat klasifikasi	
	Data training	Data testing
2 ciri	63,3	81,7
3 ciri	63,3	81,7
4 ciri	66,7	83,3
5 ciri	63,3	81,7

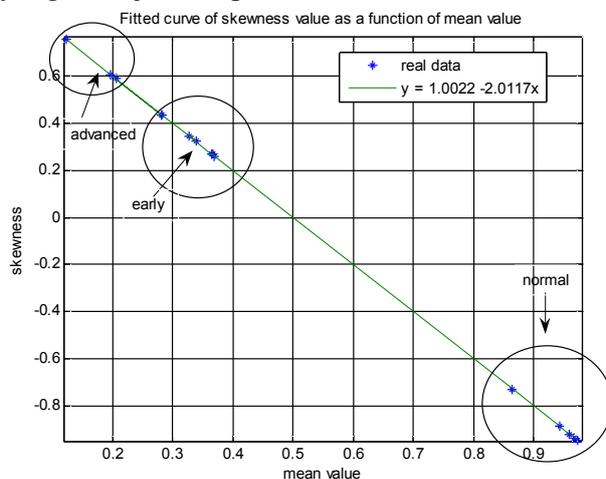
Hasil ekstraksi ciri statistik seluruh citra termogram ditunjukkan pada Tabel 2 berikut.

Tabel 2 Hasil Ekstraksi Ciri Citra Termogram

1	148.78	16.89	0.900	-22.8	104921
1	150.53	16.69	0.908	-42.1	256518
1	152.51	16.98	0.908	-54.6	377054
1	158.64	11.55	0.909	-64.6	237124
1	157.79	9.19	0.915	-58.6	123609
1	154.70	12.62	0.909	-57.2	242465
1	151.13	12.27	0.911	-24.2	40473
1	155.16	9.29	0.911	-22.4	18487
1	154.94	11.97	0.909	-51.6	189201
1	150.20	13.23	0.903	-20.7	42746
1	131.67	18.49	0.654	3.55	77599
1	136.57	21.50	0.753	-16.8	225271
1	146.32	14.99	0.887	-13.3	52784
1	145.95	13.64	0.899	-9.31	28547
1	143.77	15.64	0.898	-16.6	57747
1	144.44	16.72	0.857	-15	47895
1	139.20	17.34	0.827	-6.78	76313
1	152.22	11.73	0.915	-34.5	100765
1	147.14	18.66	0.849	-26.6	262276
1	143.48	17.18	0.891	-4.78	86248
2	123.10	27.72	0.781	-10.7	762698
2	107.60	25.68	0.728	-13.4	696188
2	116.12	28.85	0.767	-20.9	978276
2	118.84	27.49	0.257	14.33	813429
2	119.63	31.18	0.859	-25.5	1577706
2	109.81	22.40	0.871	-29.5	561711
2	127.28	25.67	0.741	-8.53	599522
2	128.13	30.09	0.787	-28.4	1275739
2	122.06	20.68	0.768	2.90	255216
2	123.81	22.41	0.810	-12.1	351429
2	131.74	25.24	0.809	-15.9	552208
2	110.80	16.14	0.660	16.82	134763
2	112.66	20.94	0.684	-4.07	408088
2	118.08	29.55	0.468	-14.7	886388
2	105.01	31.99	0.467	-34.1	1140847
2	121.88	24.02	0.833	-19.3	491594
2	121.34	22.66	0.700	-10.9	470233
2	129.90	22.20	0.891	-21.6	385435
2	128.57	24.38	0.754	2.57	384091
2	121.47	21.17	0.677	22.9	239871
3	134.43	31.82	0.425	26.9	744984
3	125.35	22.05	0.620	21.2	210763
3	123.25	26.72	0.648	-8.74	780396
3	132.53	24.57	0.719	-3.25	284585
3	124.43	29.13	0.762	-23.4	860429
3	132.80	29.06	0.534	15.63	507858
3	125.21	22.89	0.683	-3.96	373094
3	131.95	26.36	0.631	16.87	360801
3	124.86	20.39	0.735	13.35	110333
3	127.44	25.39	0.840	-13.5	548488
3	125.51	22.29	0.542	14.76	202588
3	113.05	30.95	0.537	1.58	1142268
3	118.74	21.20	0.685	16.41	370110
3	130.99	23.08	0.352	13.57	221162
3	116.78	20.91	0.758	-16.5	388024
3	122.66	21.88	0.604	25.2	225670
3	121.90	36.09	0.670	-18.1	1654430
3	127.16	22.53	0.416	26.9	283727
3	123.69	33.78	0.626	0.39	1653022
3	123.02	31.01	0.752	-19.9	1221656

Berdasarkan Tabel 2 diatas diketahui bahwa tingkat akurasi tertinggi yang diperoleh pada data *training* (pelatihan) terdapat pada klasifikasi menggunakan 4 ciri statistik sebesar 66,7% dan data *testing* (pengujian) sebesar 83,3%.

Pemanfaatan analisis diskriminan dan analisis regresi dapat digunakan secara bersamaan dengan tujuan yang berbeda. Analisis regresi digunakan untuk mencari variabel-variabel prediktor yang berpengaruh terhadap variabel respon. Selanjutnya, variabel-variabel tersebut digunakan untuk membentuk sebuah fungsi pembeda (diskriminan) yang mampu mengenali perbedaan antar kelompok. Sedangkan analisis regresi menunjukkan regresi yang linier pada hasil klasifikasi dua ciri termogram (normal, dini/early, lanjut/advanced) yaitu mean dan skewness yang ditunjukkan pada Gambar 2 berikut.



Gbr. 2 Plot hasil analisis regresi linier 2 ciri termogram mean dan skewness

V. KESIMPULAN

Pada penelitian sebelumnya telah dilakukan klasifikasi citra termogram menggunakan jaringan syaraf tiruan jenis *learning vector quantification* yang memberikan tingkat akurasi sebesar 78 %. Sedangkan dari hasil penelitian ini mampu memberikan tingkat akurasi yang lebih tinggi sebesar 83,3 % pada sampel citra termogram yang sama. Sedangkan dari hasil analisis

regresi mampu memisahkan ketiga jenis termogram, sehingga metode yang ditawarkan yaitu analisis diskriminan dan analisis regresi merupakan metode klasifikasi pengenalan pola citra termogram memberikan tingkat akurasi yang baik.

Saran : Penelitian selanjutnya perlu dilakukan klasifikasi menggunakan metode Model Markov Tersembunyi (*Hidden Markov Model*) dan Logika Fuzzy untuk melihat kinerja jaringan dan membandingkan metode yang memiliki tingkat akurasi tertinggi untuk digunakan sebagai pengenalan pola (klasifikasi) yang paling baik sebagai analisis, deteksi, atau diagnosis citra termogram.

REFERENSI

- [1] Castleman, K.R., 1996, Digital Image Processing, Prentice-Hall, Inc., Englewoods Cliff, New Jersey.
- [2] Gonzalez, R.C & Rafael E.W, 2008, Digital Image Processing, Prentice-Hall, Inc., United State, America.
- [3] Munir, R.,2004, Pengolahan Citra Digital dengan Pendekatan Logaritmik, Penerbit Informatika, Bandung.
- [4] Nurhayati, O.D, Th.Sri W, A.Susanto, Maesadji T , Jurnal Rekayasa dan Teknologi Elektro Electrician, “Peningkatan Citra Termogram untuk Klasifikasi Kanker Payudara berbasis Adaptive Neuro Fuzzy Inference System (ANFIS)”, Vol.4,No.1, Januari 2010, ISSN 1978-6042, Jurusan Teknik Elektro Fakultas Teknik, Universitas Lampung.
- [5] Nurhayati, O.D, Th.Sri W, A.Susanto, Maesadji T, The Journal of Information Technology and Electrical Engineering (JITEE), “Segmentasi Citra untuk Analisis Termogram Payudara”, Desember 2009, Vol.1 No.2.ISSN : 2085–7314, Fakultas Teknik UGM Yogyakarta.

- [6] Ovechkin, 2003, Infrared Thermography as a Method to Diagnose of the Breast Cancer, International medical online journal, 2(1):1-16, ISSN 1684-2989.
- [7] Santosa, B.,2007, Data Mining Terapan dengan Matlab, Graha Ilmu, Yogyakarta.