

KLASIFIKASI CITRA RETINA MENGGUNAKAN K-NEAREST NEIGHBOR UNTUK MENDETEKSI MAKULOPATI DIABETIK

Wahyudi Setiawan^{1*}, Fitri Damayanti²

^{1,2}Sistem Informasi, Universitas Trunojoyo

Jl. Raya Telang, PO BOX 2, Kamal, Bangkalan, Jawa Timur 69162

email : ¹wsetiawan@trunojoyo.ac.id

ABSTRAK

Pengenalan pola bertujuan untuk mengelompokkan citra ke dalam kelas tertentu berdasarkan pada ciri-ciri utama yang dimiliki. Secara umum, pengenalan pola terdiri dari 4 tahapan yaitu preprosesing, segmentasi, ekstraksi fitur dan klasifikasi. Pada penelitian ini pengenalan pola diterapkan pada deteksi tingkat Makulopati Diabetik. Makulopati Diabetik merupakan kelainan pada mata yang disebabkan oleh rusaknya pembuluh darah akibat komplikasi penyakit diabetes melitus yang terjadi di sekitar makula. Untuk mendeteksi awal penyakit makulopati diperlukan analisis dokter dari citra fundus. Citra fundus merupakan citra hasil foto retina menggunakan kamera fundus. Dataset yang digunakan yaitu MESSIDOR sebanyak 75 citra retina, terdiri dari 3 kelas yaitu Edema Makula (EM) tingkat 1, EM tingkat 2 dan EM tingkat 3. Ekstraksi fitur menggunakan Two Dimensional Linear Discriminant Analysis, sedangkan klasifikasi menggunakan k-Nearest Neighbor. Dari hasil ujicoba didapatkan prosentase pengenalan maksimal hingga 93,33%.

Kata kunci: Pengenalan Pola, Klasifikasi, Makulopati Diabetika, Two Dimensional Linear Discriminant Analysis, k-Nearest Neighbor.

ABSTRACT

Pattern recognition aims to classify the images into the certain class based on the main features. Generally, the pattern recognition consists of four stages i.e preprocessing, segmentation, feature extraction and classification. In this study, the pattern recognition applied to the detection of diabetic maculopathy. Diabetic maculopathy is a damage of blood vessels that caused by complications of diabetes mellitus in makula area. To detect early maculopathy disease, fundus image expert analysis is required. Fundus image is the image of the retinal photographs using a fundus camera. The dataset used is MESSIDOR as many as 75 images of the retina, consisting of three classes i.e macular edema (ME) grade 1, ME grade 2 and ME grade 3. Feature extraction using Two-Dimensional Linear Discriminant Analysis, while the classification using k-Nearest Neighbor. From the test results obtained a percentage of the introduction of a maximum up to 93,33%.

Keywords : *Pattern Recognition, Classification, Diabetic Retinopathy, Two-Dimensional Linear Discriminant Analysis, k-Nearest Neighbor.*

PENDAHULUAN

Pengenalan pola merupakan bagian dari ilmu pengolahan citra yang bertujuan untuk melakukan klasifikasi citra ke dalam kelas-kelas tertentu. Salah satu penerapan pengenalan pola adalah klasifikasi citra medis. Selama lebih dari 2 dekade penggunaan citra medis untuk mendiagnosis penyakit pada pasien telah umum digunakan. Citra medis biasanya diaplikasikan pada citra hasil foto rontgen untuk mengetahui kesehatan paru-paru, foto mammografi untuk

skinning kanker payudara, foto tomografi untuk mengetahui kerusakan jaringan atau kerusakan organ tertentu. Pada penelitian ini dibahas tentang citra medis fundus, yaitu berupa citra hasil foto retina untuk mendiagnosis penyakit makulopati diabetik.

Makulopati diabetik (MD) merupakan kerusakan pada pembuluh darah di bagian tengah retina (makula) yang disebabkan oleh komplikasi diabetes mellitus. MD sering terjadi pada penderita diabetes yang disertai dengan

penyakit hipertensi. Pendeteksian seseorang apakah mengalami MD, dilakukan dengan cara tes angiografi fluorescein (AF). AF merupakan cara untuk mendeteksi terjadinya kerusakan pada pembuluh darah pada retina dan berapa banyaknya terjadi kerusakan tersebut. Tes selain AF yaitu Optic Coherence Tomography (OCT). Proses OCT menggunakan kamera fundus untuk mendapatkan citra retina. Tujuannya untuk mengetahui kerusakan pada pembuluh darah retina (Hammes et al., 2014).

Pengujian AF dan OCT memerlukan waktu relatif lama. Diperlukan diagnosis dan analisis dari pakar untuk mengetahui kerusakan retina dan tergolong kerusakan edema makula pada tingkat 1, 2 atau 3 (Wu et al., 2013).

Penelitian ini menggunakan dataset MESSIDOR. MESSIDOR merupakan dataset citra fundus yang dapat diunduh secara online (<http://www.adcis.net/en/DownloadThirdParty/Messidor.html>). MESSIDOR dapat digunakan untuk mendeteksi tingkat retinopati diabetik dan tingkat edema makula yang mengakibatkan penyakit makulopati diabetik. Tingkat edema makula ditentukan oleh hard exudates (HE). HE merupakan infiltrasi lipid ke dalam retina, bentuknya tidak beraturan dan berwarna kekuningan. Citra yang digunakan sebanyak 75 citra terdiri dari 3 kelas, masing-masing kelas terdiri dari 25 citra (Decenciere et al., 2014). Berikut ini merupakan klasifikasi tingkat edema makula :

0 : Mild

1 : Moderate

2 : Severe

METODE

Pada penelitian ini dibuat sebuah sistem tingkat makulopati diabetik. Sistem terdiri dari empat tahapan yaitu preprosesing, segmentasi, ekstraksi fitur dan klasifikasi. Sistem terdiri dari 2 (dua) bagian yaitu pelatihan dan pengujian. Pada bagian pelatihan, sistem memerlukan input sejumlah data citra yang akan digunakan untuk mengenali ciri-ciri utama dari 3 kelas yang berbeda. Selanjutnya citra yang digunakan sebagai data uji dilakukan pencocokan dengan data pelatihan. Ciri-ciri utama menentukan sebuah citra akan terklasifikasi ke dalam kelas tertentu.

Preprocessing

Proses pengolahan awal bertujuan untuk memperbaiki citra yang digunakan sebagai data

ujicoba. Beberapa tahapan dilakukan untuk perbaikan citra, diantaranya yaitu citra RGB diubah menjadi citra aras keabuan green channel, Histogram CLAHE dan filter gaussian.

Green channel

Sebuah citra RGB terdiri dari citra Red, Green dan Blue. Citra dengan aras keabuan biasanya membagi warna citra menjadi Red, Green dan Blue dengan bobot nilai yang sama yaitu 0,33.

$$I(x, y) = \alpha.R + \beta.G + \gamma.B \quad (1)$$

$I(x,y)$ merupakan skala keabuan yang ditentukan oleh nilai α , β , γ . Pada skala keabuan biasa, nilai α , β , γ sama yaitu 0,33. Sedangkan untuk aras keabuan green channel nilai β diberi nilai 1, sedangkan α , γ diberi nilai 0. Refleksi cahaya Green Channel lebih baik dibandingkan aras keabuan yang lain (Kolar and Harabis, 2009).

Ekualisasi Histogram

Ekualisasi histogram melakukan penyesuaian nilai piksel untuk meningkatkan kontras pada citra. Ekualisasi histogram melakukan perataan dari nilai keabuan suatu citra. Fungsi Cummulative Distribution Function (CDF). CDF dihitung menggunakan persamaan 2.

$$f(k) = \frac{(N-1)}{M} \sum_{k=0}^n h(k); \quad (2)$$

$$n = 1,2,3,\dots, N-1$$

M = piksel

N = grayscale

H(k) = histogram suatu nilai k

Adaptive Histogram Equalization (AHE) merupakan teknik perbaikan pada kontras lokal citra. AHE menggunakan data lokal citra. Lokal citra disebut sebagai region size. Dengan ukuran piksel citra tertentu, dibuat region size. Citra lokal terdiri dari 3 jenis yaitu Corner Region (CR), Inner Region (IR) dan Border Region (BR) (Kanditami dkk, 2014).

Corner Region merupakan piksel pojok pada suatu citra lokal. Border Region (BR) merupakan piksel paling luar selain CR. Inner Region (IR) merupakan piksel yang terletak di

dalam suatu lokal citra. Untuk memberikan gambaran yang jelas tentang Region Size dapat dilihat pada gambar 1.

CR	BR	BR	CR
BR	IR	IR	BR
BR	IR	IR	BR
CR	BR	BR	CR

Gambar 1. Struktur Region Size

Cara untuk mendapatkan nilai gray level pada setiap piksel yaitu dengan mendapatkan nilai Cumulative Distribution Function (CDF). CDF didapatkan sesuai dengan persamaan 2. Untuk mengatasi nilai yang berlebih pada AHE maka digunakan CLAHE. CLAHE menggunakan nilai batas untuk histogram. Nilai batas disebut sebagai clip limit, merupakan nilai batas tertinggi suatu histogram. Clip limit dapat dihitung menggunakan persamaan 3.

$$\beta = \frac{M}{N} \left(1 + \frac{\alpha}{100} (S_{max} - 1)\right) \quad (3)$$

M = luas region size

N = nilai grayscale (256)

α = clip factor, batas limit antara 0 – 100

Histogram memiliki nilai lebih (excess) sehingga kelebihan nilai akan disimpan di area clip limit sehingga didapatkan nilai grayscale lebih merata.

Filter Gaussian

Filter bertujuan untuk menghilangkan noise yang terdapat di citra. Filter Gaussian merupakan filter dengan fungsi gaussian untuk pembobotan anggotanya. Filter ini sangat baik untuk menghilangkan derau yang bersifat sebaran normal. Secara alami derau juga memiliki sebaran *Gaussian*, sehingga secara teoritis akan menjadi netral jika dilawan dengan fungsi lain yang juga memiliki fungsi *Gaussian*, hal ini disebut sebagai *zero mean*. *Zero mean* dari fungsi *Gaussian* dengan nilai pembobotan 2 dimensi ditunjukkan pada persamaan (4) (Ahmad, 2005).

$$\frac{g(x,y)}{k} = e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (4)$$

dengan k = konstanta normalisasi dan σ menyatakan standar deviasi dari distribusi. Fungsi diatas diasumsikan memiliki *zero mean* (pusat distribusi pada garis $x=0$). Semakin besar nilai σ maka kurva distribusi *Gaussian* semakin Bentuk 2D dari fungsi gaussian adalah :

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (5)$$

Tabel 1. Tapis distribusi *Gaussian* 2-D dengan ukuran tapis 5 x 5

(x, y)	-2	-1	0	1	2
-2	0,00 29	0,01 31	0,02 15	0,01 31	0,00 29
1	0,01 31	0,05 85	0,09 65	0,05 85	0,01 31
0	0,02 15	0,09 65	0,15 92	0,09 65	0,02 15
1	0,01 31	0,05 85	0,09 65	0,05 85	0,01 31
2	0,00 29	0,01 31	0,02 15	0,01 31	0,00 29

Segmentasi

Segmentasi merupakan teknik yang digunakan untuk memberi pola terhadap region yang memiliki kemiripan. Segmentasi juga bertujuan untuk memisahkan obyek asli dengan background. Segmentasi menggunakan Penandaan komponen terhubung dan masking.

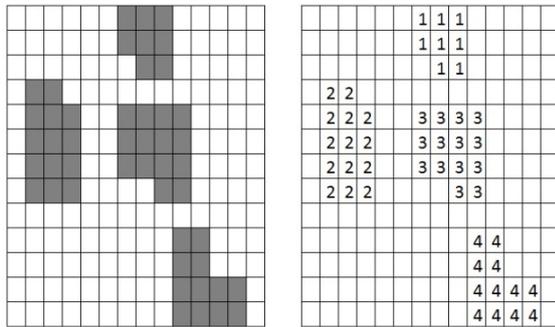
Penandaan komponen terhubung (*connected component labelling*)

Penandaan komponen terhubung merupakan sekumpulan piksel yang terhubung antara satu piksel dengan piksel lainnya. Piksel-piksel yang letaknya berdekatan dan memiliki kemiripan akan diberi tanda berupa angka numerik 1, 2, 3 dan seterusnya untuk menunjukkan bahwa kumpulan piksel tersebut memiliki kemiripan.

Beberapa jenis algoritma dapat digunakan untuk menyelesaikan permasalahan ini diantaranya Algoritma rekursif, algoritma sequential dan region boundary. Pada penelitian ini digunakan algoritma rekursif, sebagai berikut (Jain, 2015):

1. Scan citra untuk menemukan piksel bernilai bukan nol dan beri tanda label baru, misalnya dimulai dengan angka 1
2. Periksa piksel di sekelilingnya yang terhubung dengan piksel tersebut, tandai dengan angka yang sama

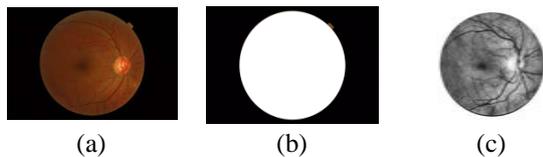
3. Stop, jika tidak ditemukan lagi piksel dengan unlabeled 0
4. Ulangi ke langkah 1, jika ditemukan piksel yang bukan nol, beri angka 2 dan seterusnya Untuk memberikan gambaran lebih jelas terhadap algoritma rekursif. Gambar 2 menunjukkan algoritma tersebut.



Gambar 2. Penandaan komponen terhubung

Masking

Masking bertujuan untuk memisahkan antara obyek citra asli dengan citra background. Salah satu teknis untuk melakukan masking yaitu dengan melakukan operasi subtracting antara citra asli dengan citra masking (Gonzales and Woods, 2012). Pada gambar 3 c. merupakan hasil masking.



Gambar 3. (a) citra asli, (b) citra masking, (c) citra hasil preprosesing dan masking

Ekstraksi ciri

Ekstraksi ciri merupakan proses untuk mendapatkan ciri-ciri utama yang dapat membedakan antara citra satu dengan citra lainnya. Selain itu, ekstraksi ciri juga berfungsi sebagai reduksi dimensi.

Reduksi dimensi perlu dilakukan, karena data ujicoba memiliki ukuran yang besar, yaitu 2240x1488 piksel. Dari jumlah citra pelatihan sebanyak 30, maka dimensi awal adalah 2240x1488 = 3.333.120. Citra akan direduksi dimensi menjadi 80, yaitu sebanyak data pelatihan.

Ekstraksi ciri yang digunakan yaitu *Two-Dimensional Linear Discriminant Analysis (2DLDA)*

Two-Dimensional Linear Discriminant Analysis

2DLDA merupakan pengembangan dari LDA. Jika pada LDA, matriks diubah menjadi vektor, maka pada 2DLDA yang digunakan untuk proses algoritmanya tetap berupa matriks. Pada dasarnya LDA bertujuan untuk mendapatkan transformasi proyeksi optimal berupa reduksi dimensi, serta memaksimalkan matriks kovarian antar kelas S_w serta meminimalkan matriks kovarian dalam kelas S_b . Metode 2DLDA mengatasi *singular problem* jika terjadi penerapan kasus pada metode LDA.

Berikut ini merupakan algoritma 2DLDA (Ye et al., 2004) :

Input : $A_1, \dots, A_n, \ell_1, \ell_2$

Output : L, R, B_1, \dots, B_n

Hitung mean M_i dari tiap kelas menggunakan persamaan

$$M_i = \frac{1}{n_i} \sum x \in \pi_i X; \tag{6}$$

Hitung mean Global M

$$M = \frac{1}{n} \sum_{i=1}^k \sum x \in \pi_i X; \tag{7}$$

$$R_0 \leftarrow (I \ell_2, 0)^T \tag{8}$$

For j from 1 to I

$$S_w^R \leftarrow \sum_{i=1}^k \sum x \in \pi_i (X - M_i) R_{j-1} R_{j-1}^T (X - M_i)^T \tag{9}$$

$$S_w^R \leftarrow \sum_{i=1}^k \sum x \in \pi_i (X - M_i) R_{j-1} R_{j-1}^T (X - M_i)^T \tag{10}$$

Hitung vektor eigen pertama $\ell_1 \{\phi_l^L\}_{l=1}^{\ell_1}$ dari $(S_w^R)^{-1} S_b^R$ (11)

$$L_j \leftarrow [\phi_1^L, \dots, \phi_{\ell_1}^L] \tag{12}$$

$$S_w^L \leftarrow \sum_{i=1}^k \sum x \in \pi_i (X - M_i) L_{j-1} L_{j-1}^T (X - M_i)^T \tag{13}$$

$$S_w^L \leftarrow \sum_{i=1}^k \sum x \in \pi_i (X - M_i) L_{j-1} L_{j-1}^T (X - M_i)^T \tag{14}$$

Hitung vektor eigen pertama $\ell_2 \{\phi_l^R\}_{l=1}^{\ell_2}$ dari $(S_w^L)^{-1} S_b^L$ (15)

$$R_j \leftarrow [\phi_1^R, \dots, \phi_{\ell_2}^R] \tag{16}$$

End For

$$L \leftarrow L_I, R \leftarrow R_I \tag{17}$$

$$B_L \leftarrow L^T A_\ell R, \text{ for } \ell = 1, \dots, n; \tag{18}$$

Return (L,R,B1,...,Bn)

Tabel 2 merupakan penjelasan notasi dari algoritma 2DLDA.

Notasi	Deskripsi
n	jumlah citra pada dataset
k	jumlah kelas pada dataset
A _i	matriks citra ke i
A _i	vektor citra ke i
r	jumlah baris pada citra ai
c	jumlah kolom pada citra ai
N	dimensi ai (n= r*c)
π _j	kelas ke j pada dataset
L	transformasi matriks (kiri) 2DLDA
R	transformasi matriks (kanan) 2DLDA
I	
B _i	jumlah iterasi pada 2DLDA reduksi pada A _i dengan 2DLDA
ℓ ₁	jumlah baris pada B _i
ℓ ₂	jumlah kolom pada B _i

Klasifikasi k-Nearest Neighbor menggunakan jarak Euclidian Distance

k-Nearest Neighbor merupakan salah satu jenis klasifikasi sederhana yang dapat digunakan untuk mengelompokkan citra ke dalam kelas-kelas tertentu. Prinsip kerja dari algoritma k-NN adalah menghitung jarak dari citra yang digunakan untuk data pelatihan dengan citra uji (Lammertsma, 2015).

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (19)$$

dimana :

d = jarak

i = variabel data

p = dimensi data

x_{2i} = data pelatihan

x_{1i} = data uji

HASIL DAN PEMBAHASAN

Berbagai macam skenario diujicobakan pada citra makula edema. Pada tabel 3. Dijelaskan tentang skenario ujicoba tersebut.

Tabel 2. Skenario Ujicoba

Jumlah data training	Jumlah data uji	Akurasi optimal
30	45	75,56%
45	30	80%
60	15	93,33%

SIMPULAN DAN SARAN

Presentase keberhasilan akurasi paling maksimal sebesar 93,33%. Hal ini dipengaruhi oleh semakin banyaknya data pelatihan maka semakin baik presentase keakuratan.

Pada penelitian selanjutnya akan dilakukan skenario pengujian dengan jumlah data uji tetap data latih yang berbeda. Selain itu, metode ekstraksi ciri dapat dikembangkan menggunakan metode statistik regresi non parametrik.

UCAPAN TERIMA KASIH

Penelitian ini didanai oleh DRPM DIKTI. Penelitian ini merupakan penelitian tahun ke 2 dari 2 tahun rencana penelitian. Penulis mengucapkan banyak terima kasih kepada DRPM DIKTI yang telah memberikan dukungan dana hingga selesainya penelitian.

DAFTAR PUSTAKA

- Ahmad, U., 2005. *Pengolahan Citra Digital dan Teknik Pemrogramannya*, Graha Ilmu, Yogyakarta.
- Decenciere, E., Zhang, X., Cazugule, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez-Varela, J., Massin, P., Erginay, A., Charton, B. and Klein, J, 2014, *Feedback on a Publicity Distributed Image Database: The Messidor Database*, Image Anal Stereol, 33, pp:231-234
- Gonzalez, R.C., Woods, R.E., 2012. *Digital image processing*, second edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.
- Hammes, H.P., Lemmen, K.D. and Bertram, B., 2014, *Diabetic Retinopathy and Maculopathy*, Exp Clin Endocrinol Diabetes, 122, pp: 387-390.
- Jain, R. 2015, *Connected Component*, [ONLINE], Available at <https://www.cse.unr.edu/~bebis/CS791E/Notes/ConnectedComponents.pdf>
- Kanditami, F., Saepudin, D., dan Rizal, A, 2014, *Analisis CLAHE dan Region Growing dalam Deteksi Gejala Kanker*

- payudara pada Citra Mammogram*,
Jurnal Elektro, Vol 7 No 1, hal 15-28
- Kolar, R. and Harabis, V., 2009. *Automatic Rigid Registration and Analysis of Colour Fundus Image in Patient with Diabetic Retinopathy*, IFMBE proceedings, vol 25/11, 251-24.
- Lammertsma, P., 2015, *k-Nearest Neighbor*, [ONLINE], available at paul.luminos.nl/download/document/knn.pdf.
- Wu, L., Fernandez-oaiza, P., Sauma, J., Hernandez-Bogantes, W., and Masis, M. 2013, *Classification of Diabetic Retinopathy and Diabetic Macular Edema*, World J Diabetes, 4(6), pp: 290-294.
- Ye, J., Janardan, R. and Li, Q., 2004, *Two Dimensional Linear Discriminant Analysis*, [ONLINE], available at <https://www.researchgate.net/publication/s/221619249>