

PERANCANGAN SISTEM PENDETEKSI LEUKOPLAKIA MELALUI CITRA KLINIS RONGGA MULUT BERBASIS GRAY LEVEL CO-OCCURRENCE MATRIX DAN K-NEAREST NEIGHBOR

DESIGN OF LEUKOPLAKIA DETECTION SYSTEM THROUGH CLINICAL PICTURE OF ORAL BASED ON GRAY LEVEL CO-OCCURRENCE MATRIX AND K-NEAREST NEIGHBOR

I Wayan Agus Sugiarsa^[1]

Iwan Iwut T.^[2]

Hilman Fauzi T. S. P.^[3]

Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

wayanagussugik@gmail.com

Abstrak

Homogeneous Leukoplakia, Hairy Leukoplakia dan Oral Candidiasis merupakan lesi putih yang ditemukan pada rongga mulut. *Homogeneous Leukoplakia* adalah lesi praganas yang disebabkan oleh iritasi kronis. *Hairy Leukoplakia* merupakan salah satu bentuk *Leukoplakia*, hanya saja *Hairy Leukoplakia* tidak termasuk lesi praganas. *Oral Candidiasis* merupakan jamur yang tumbuh pada rongga mulut, memiliki gambaran klinis yang hampir sama dengan *Leukoplakia*. Sejak *Hairy Leukoplakia* menjadi ciri utama yang ditemukan pada pasien *AIDS*, perbedaannya dengan lesi lain terutama *Leukoplakia* yang memiliki citra klinis yang mirip dengan *Hairy Leukoplakia* menjadi sangat penting. Oleh karena itu diperlukan suatu sistem yang dapat mengidentifikasi dan membedakan ketiga lesi tersebut dengan akurasi yang maksimal. Ketiga jenis lesi putih tersebut bisa dideteksi dengan mengolah informasi tekstur yang terdapat di daerah-daerah tertentu pada rongga mulut.

Perkembangan teknologi pengolahan citra digital dapat dijadikan solusi untuk mengurangi ketidakakuratan penglihatan manusia. Pada tugas akhir ini dibuat suatu sistem yang bisa mengidentifikasi citra lesi *Homogeneous Leukoplakia, Hairy Leukoplakia dan Oral Candidiasis* berdasarkan informasi tekstur citra dengan metode ekstraksi ciri *Gray Level Co-Occurrence Matrix (GLCM)* dan metode klasifikasi *K-Nearest Neighbor (K-NN)*. Citra didapatkan dari sumber resmi berupa citra digital. Untuk mengetahui performansi sistem, diukur pula tingkat akurasi sistem. Data diolah dengan menggunakan *software MATLAB R2014a*.

Setelah dilakukan pengujian, didapat kesimpulan bahwa sistem dapat membedakan *Homogeneous Leukoplakia, Hairy Leukoplakia dan Oral Candidiasis* dengan akurasi tertinggi 84,848%. Kombinasi parameter yang memberikan akurasi tertinggi yaitu sudut *GLCM 90°*, *property GLCM Difference Entropy*, nilai variabel *K 1* dan *distance K-NN Euclidean*.

Kata Kunci: Kata kunci: *Homogeneous Leukoplakia, Hairy Leukoplakia, Oral Candidiasis, Gray Level Co-Occurrence Matrix (GLCM), K-Nearest Neighbor (K-NN)*.

Abstract

Homogeneous Leukoplakia, Hairy Leukoplakia, and Oral Candidiasis is a white lesion found in the Oral cavity. *Homogeneous Leukoplakia* is a pre-cancerous lesions caused by chronic irritation. *Hairy Leukoplakia* is one form of *Leukoplakia*, just *Hairy Leukoplakia* does not include pre-cancerous lesions. *Oral Candidiasis* is a fungus that grows in the mouth, has a clinical picture similar to *Leukoplakia*. Since *Hairy Leukoplakia* became the main characteristic found in *AIDS* patients, the Difference with the others mainly *Leukoplakia* lesions that have similar clinical image with *Hairy Leukoplakia* becomes very important. Therefore we need a sistem that can identify and distinGUIsh the three lesions with maximum accuracy. Three types of white lesions can be detected by processing the texture information contained in certain areas of the Oral cavity.

The development of digital image processing technology can be used as a solution to reduce the inaccuracies of human vision. In this final project created a sistem that could identify the image of the lesions *Homogeneous Leukoplakia, Hairy Leukoplakia, and Oral Candidiasis* based image texture information, by using a feature extraction method *Gray Level Co-Occurrence Matrix (GLCM)* and using the classification *K-Nearest Neighbor (K-NN)*. Image obtained from official sources in the form of a digital image. To determine the

performance of the sistem, measured the level of accuracy of the sistem. The data is processed by using software MATLAB R2014a.

After testing, concluded that the sistem can distinguish Homogeneous Leukoplakia, Hairy Leukoplakia, and Oral Candidiasis with the highest accuracy of 84.848%. The combination of parameters that provide the highest accuracy GLCM angle 90° , GLCM Difference Entropy property, the value of K 1 and K-NN Euclidean distance.

Keywords: Homogeneous Leukoplakia, Hairy Leukoplakia, Oral Candidiasis, Gray Level Co-Occurrence Matrix (GLCM), K-Nearest Neighbor (K-NN).

1. Pendahuluan

Homogeneous Leukoplakia adalah lesi pra-kanker yang ditemukan pada rongga mulut. *Hairy Leukoplakia* adalah lesi yang menyerupai *Homogeneous Leukoplakia* namun bukan lesi pra-kanker. *Oral Candidiasis* adalah jamur yang ditemukan pada rongga mulut menyerupai *Homogeneous Leukoplakia*. Ketiga lesi tersebut memiliki ciri yang sama namun penyebab dan pengobatannya berbeda-beda. Dalam ilmu pengolahan citra digital, satu jenis tekstur dapat dibedakan antara satu dengan yang lainnya dengan memanfaatkan metode ekstraksi ciri dan klasifikasi.

Homogeneous Leukoplakia, *Hairy Leukoplakia* dan *Oral Candidiasis* dapat dibedakan berdasarkan teksturnya. Oleh karena itu dengan menggunakan metode pengolahan citra digital, ketiga lesi tersebut dapat diklasifikasikan berdasarkan tekstur masing-masing lesi.

Pada penelitian sebelumnya [1] *Homogeneous Leukoplakia* dan *Hairy Leukoplakia* telah diteliti perbedaannya namun belum menggunakan metode pengolahan citra digital. *GLCM* merupakan metode ekstraksi ciri yang bisa digunakan untuk mengekstrak informasi suatu citra [4]. *K-NN* adalah metode yang bisa digunakan untuk klasifikasi suatu fitur citra [8].

Pada tugas akhir ini dirancang suatu sistem yang mempunyai kemampuan untuk membedakan *Homogeneous Leukoplakia*, *Hairy Leukoplakia* dan *Oral Candidiasis* pada rongga mulut melalui foto citra rongga mulut dengan metode ekstraksi ciri *Gray Level Co-Occurrence Matrix (GLCM)* dan metode klasifikasi *K-Nearest Neighbor (K-NN)*. Judul dari Tugas Akhir ini adalah "Perancangan Sistem Pendeteksi *Leukoplakia* Melalui Citra Klinis Rongga Mulut Berbasis *Gray Level Co-occurrence Matrix* dan *K-Nearest Neighbor*".

Pada metode ekstraksi ciri *GLCM* akan diteliti pengaruh sudut dan *property* terhadap akurasi sistem. Pada metode klasifikasi *K-NN* akan diteliti pengaruh nilai K dan *distance* terhadap akurasi sistem. Setelah didapatkan kombinasi ke empat parameter tersebut yang dapat memberikan akurasi yang maksimal, akan dihitung pula waktu komputasi untuk enam kombinasi terbaik sehingga performansi sistem dapat diketahui.

Data citra digital yang diolah adalah citra *RGB* dengan *file* bertipe *.jpg dengan ukuran citra 100 x 200 pixel. Perancangan sistem menggunakan *software* Matlab R2014a. Citra inputan adalah citra rongga mulut yang terindikasi *Homogeneous Leukoplakia*, *Hairy Leukoplakia* dan *Oral Candidiasis*. Sistem bekerja secara *non-realtime*.

2. Dasar Teori

2.1 Gambaran Sistem

Secara umum terdapat dua tahap utama yaitu pelatihan sistem dan pengujian sistem. Pada tahap pelatihan sistem dilakukan pelatihan terhadap data latih. Pada tahap pengujian sistem dilakukan pengujian terhadap data uji. Keluaran sistem adalah hasil klasifikasi citra uji. Pada tahap *Pre-processing* dilakukan *cropping*, konversi citra *RGB* menjadi citra *grayscale*, dan equalisasi *Histogram*. *Pre-processing* diterapkan pada tahap pelatihan dan pengujian. Pada tahap pelatihan sistem terdapat dua proses yaitu proses *Pre-processing* dan ekstraksi ciri. Pada tahap ekstraksi ciri digunakan metode *GLCM*. Hasil ekstraksi ciri disimpan dalam database dalam bentuk file .mat. Pada tahap pengujian sistem terdapat tiga proses yaitu proses *Pre-processing*, ekstraksi ciri dan klasifikasi. Pada tahap ekstraksi ciri digunakan metode *GLCM*. Pada tahap klasifikasi dilakukan klasifikasi data uji menggunakan metode *K-NN*. Keluaran sistem didapatkan setelah proses klasifikasi.

Proses pengujian sistem dilakukan setelah proses pelatihan. Pada proses pelatihan dibuat data latih yang nantinya diperlukan dalam tahap pengujian yaitu tahap klasifikasi data uji. Pengujian sistem dilakukan dengan membandingkan hasil output klasifikasi pada tahap pengujian sistem dengan output yang seharusnya atau target output data uji sehingga didapatkan akurasi sistem. Berikut adalah langkah-langkah pengujian sistem secara umum. Setiap parameter input maupun output pada tahap ekstraksi ciri dan klasifikasi dipasangkan sampai semua pasangan parameter diketahui akurasinya. Terakhir, waktu komputasi dihitung dan dibandingkan dari 6 pasangan parameter yang mempunyai akurasi sistem terbaik. Performansi sistem dilihat dari banyak jumlah hasil pengujian yang sesuai dengan target kelas citra (akurasi) dan waktu pengujian. Namun Parameter performansi waktu pengujian diperhitungkan setelah didapatkan susunan parameter input maupun output yang menghasilkan akurasi maksimal.

2.2 Penyakit Pada Rongga Mulut [1]

Ada banyak sekali penyakit yang ditemukan pada rongga mulut. Diantaranya adalah *Homogeneous Leukoplakia*, *Hairy Leukoplakia* dan *Oral Candidiasis*. Ketiga lesi tersebut mempunyai gambaran klinis yang mirip namun penyebab dan penanggulangannya berbeda-beda.

2.2.1 *Homogeneous Leukoplakia*

Homogeneous Leukoplakia disebut juga *Leukoplakia simpleks*, berupa lesi berwarna keputih-putihan, dengan permukaan rata, licin atau berkerut, dapat pula beralur atau berupa suatu peninggian dengan pinggiran yang jelas. Sangat besar kemungkinan *Homogeneous Leukoplakia* berkembang menjadi sel kanker jika tidak ditangani dengan segera. *Homogeneous Leukoplakia* dapat diobati dengan mengangkat atau dengan menghancurkan sel yang terinfeksi. Data *Homogeneous Leukoplakia* yang digunakan untuk data latih sebanyak 20 citra dan data uji sebanyak 24 citra.



Gambar 2.1 *Homogeneous Leukoplakia*

2.2.2 *Hairy Leukoplakia*

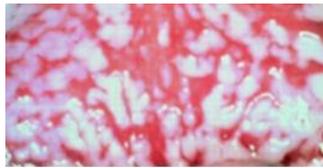
Hairy Leukoplakia merupakan lesi putih yang hampir selalu terjadi unilateral atau bilateral pada tepi lateral lidah, sering tampak menyerupai rambut atau bergelombang, dapat pula seperti plak. *Hairy Leukoplakia* biasa ditemukan pada pasien HIV AIDS. Data *Hairy Leukoplakia* yang digunakan sebagai data latih sebanyak 20 citra dan data uji sebanyak 18 citra.



Gambar 2.2 *Hairy Leukoplakia*

2.2.3 *Oral Candidiasis* [2]

Oral Candidiasis adalah jamur yang ditemukan pada rongga mulut. Pada umumnya *Oral Candidiasis* disebabkan oleh kondisi imun yang menurun, kemudian ada kasus yang disebabkan karena penggunaan alat tertentu pada rongga mulut yang tidak dibersihkan dengan benar, misalnya kawat gigi dan gigi palsu. *Oral Candidiasis* memiliki beberapa ciri klinis. Diantaranya lesi yang menampilkan membrane yang melekat longgar yang terdiri dari organism jamur dan debris cellular yang meninggalkan sebuah peradangan, terkadang area perdarahan jika pseudomembran dihilangkan. Data *Oral Candidiasis* yang digunakan sebagai data latih sebanyak 20 citra dan data uji sebanyak 24 citra.



Gambar 2.3 Oral Candidiasis

2.3 Citra Digital [3]

Citra Digital merupakan suatu fungsi kontinu dari intensitas cahaya dalam bidang dua dimensi yang diubah ke dalam bentuk diskrit atau nilai digital yang disebut dengan pixel. Format citra digital adalah berbentuk matriks N x M, dimana N menyatakan baris/tinggi dan M menyatakan kolom/lebar.

2.3.1 Jenis Citra [4]

Citra berwarna atau citra RGB adalah citra yang menyajikan warna dalam bentuk komponen R (Red), G (Green) dan B (Blue). Citra grayscale adalah citra yang menangani gradasi warna hitam dan putih yang menghasilkan efek warna abu-abu. Pada citra grayscale warna dinyatakan dengan intensitas yang berkisar antara 0 sampai dengan 255. Nilai 0 menyatakan hitam dan nilai 255 menyatakan putih.

2.3.2 Konversi Citra RGB ke Grayscale [4]

Persamaan yang digunakan untuk mengkonversi citra warna RGB ke citra grayscale adalah:

Dengan menggunakan persamaan (1) adalah intensitas warna grayscale pada pada kordinat (x,y) adalah (1) $I(x,y) = 0,2989 \times R(x,y) + 0,5870 \times G(x,y) + 0,1141 \times B(x,y)$ dengan menggunakan persamaan (2) adalah intensitas warna Red pada kordinat (x,y), $R(x,y)$ adalah intensitas warna Green pada kordinat (x,y), dan $B(x,y)$ adalah intensitas warna Blue pada kordinat (x,y). Persamaan 2. Diterapkan pada setiap pixel citra warna RGB.

2.3.3 Equalisasi Histogram [5]

Histogram Equalization adalah suatu proses perataan Histogram, dimana distribusi nilai derajat keabuan pada suatu citra dibuat rata. Untuk dapat melakukan HE ini diperlukan suatu fungsi distribusi kumulatif yang merupakan perhitungan kumulatif dari Histogram. Fungsi distribusi kumulatif dapat didefinisikan sebagai berikut.

$$H_k = \sum_{i=0}^{k-1} h(i), \quad k = 1, 2, 3, \dots, (N - 1) \tag{2}$$

M menyatakan piksel sedangkan N menyatakan grayscale, dan $h(k)$ menyatakan Histogram pada suatu nilai gray value k.

2.4 Gray Level Co-Occurrence Matrix [6]

Ekstraksi ciri statistik orde kedua merupakan metode ekstraksi ciri yang efektif digunakan untuk mengekstrak informasi yang terdapat pada suatu citra. Ekstraksi ciri statistik orde kedua mempertimbangkan hubungan antara dua piksel (piksel yang bertetangga) pada citra. Untuk kebutuhan analisisnya, analisis tekstur orde dua memerlukan bantuan matriks kookurensi (matrix co-occurrence) untuk citra keabuan, biasanya disebut GLCM. Analisa tekstur orde dua lebih baik dalam merepresentasikan tekstur citra dalam parameter-parameter terukur. Ekstraksi ciri statistik orde kedua dilakukan dengan matriks kookurensi, yaitu suatu matriks antara yang merepresentasikan hubungan ketetanggaan antar piksel dalam citra pada berbagai arah orientasi dan jarak spasial. Matriks kookurensi merupakan matriks berukuran L x L (L menyatakan banyaknya tingkat keabuan) dengan elemen-elemen yang merupakan distribusi probabilitas bersama (joint probability distribution) dari pasangan dua tingkat keabuan. Matriks kookurensi yang dibentuk disebabkan berjarak r (dengan sudut θ) yang berfokus pada kedua piksel yang dihitung dengan pengelompokan sebagai berikut:

$$P(i,j,r,\theta) = \frac{1}{N^2} \sum_{x,y} \sum_{x',y'} \delta(x-x',y-y') \delta(x-x'+r,y-y'+\theta) \tag{3}$$

$$\sqrt{\frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} \right)} \sqrt{\frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} \right)}$$

(12)

2.6 MATIAB [8]

MATLAB (*Matrix Laboratory*) adalah sebuah bahasa pemrograman untuk analisis dan komputasi numerik dan merupakan suatu bahasa pemrograman matematika lanjutan yang dibentuk dengan dasar pemikiran menggunakan sifat dan bentuk matriks. MATLAB telah berkembang menjadi sebuah *software* pemrograman yang canggih yang berisi fungsi-fungsi *built-in* untuk melakukan tugas pengolahan sinyal, aljabar linier, dan kalkulasi matematis lainnya.

3. Pembahasan

3.1 GLCM

Sudut yang memberikan akurasi tertinggi adalah 90^0 yaitu 84,85%. Sudut tersebut menghasilkan akurasi yang tinggi karena ketiga kelas uji mempunyai paling banyak perbedaan fitur *Difference Entropy* pada sudut 90^0 . Pada sudut tersebut juga terdapat paling banyak persamaan fitur citra pada satu kelas. Semakin besar atau semakin kecil sudut *GLCM* tidak mempengaruhi akurasi.

Property GLCM yang memberikan akurasi paling tinggi adalah *Difference Entropy* dengan akurasi 84,85%. *Property GLCM* tersebut menghasilkan akurasi yang tinggi karena selisih keragaman atau *Entropy* satu kelas data latih memiliki nilai yang berdekatan. Selain itu nilai *Difference Entropy* antara satu kelas dengan kelas lain juga memiliki jarak yang cukup besar. Sehingga pada tahap klasifikasi, *K-NN* dapat membedakan masing-masing kelas dengan lebih baik. Nilai *property GLCM* yang paling baik adalah nilai yang mempunyai jarak terdekat dengan kelas masing-masing dan terjauh dengan kelas lain.

3.2 K-NN

Nilai variabel *K* yang memberikan akurasi paling tinggi adalah 1 yaitu 84,85%. Nilai variabel *K* tersebut menghasilkan akurasi yang tinggi karena ketika nilai variabel *K* 1 batas antar kelas semakin jelas. Semakin besar nilai variabel *K* maka akurasi sistem cenderung semakin berkurang. Semakin tinggi nilai variabel *K* dapat mengurangi pengaruh noise pada citra namun dapat membuat batasan antar kelas menjadi kabur.

Distance Euclidean dan *Cityblock* menghasilkan akurasi tertinggi 84,85%. Pemilihan *distance K-NN* sangat mempengaruhi akurasi sistem. Akurasi sistem yang dihasilkan *distance Euclidean* dan *Cityblock* selalu sama untuk semua skenario kombinasi karena ekstraksi ciri *GLCM* hanya menghasilkan satu nilai output untuk setiap citra sehingga dimensi *K-NN* hanya 1 dimensi. Jika dimasukkan ke dalam persamaan *Euclidean* dan *Cityblock*, untuk *K-NN* dengan 1 dimensi, jarak yang dihasilkan *Euclidean* akan sama dengan jarak yang dihasilkan *Cityblock*. *Cosine* dan *Correlation* tidak efektif digunakan untuk sistem karena kedua *distance* tersebut hanya menentukan kemiripan dan tidak menentukan kelas dari data uji.

3.3 Waktu Komputasi

Kombinasi parameter yang memerlukan waktu komputasi yang paling sedikit adalah sudut *GLCM* 90^0 , *property GLCM Difference Entropy*, nilai variabel *K* 1 dan *distance K-NN Euclidean* yaitu 0,63ms. Perbedaan waktu komputasi untuk seluruh skenario sangat kecil. Seluruh skenario mempunyai performa waktu komputasi yang bagus.

4. Kesimpulan

Berdasarkan hasil pengujian dan analisis pengujian sistem pendeteksi *Leukoplakia* menggunakan citra rongga mulut, maka dapat ditarik kesimpulan sebagai berikut.

1. Sistem yang dapat membedakan *Hairy Leukoplakia*, *Homogeneous Leukoplakia* dan *Oral Candidiasis* telah berhasil dibuat dengan akurasi maksimal sebesar 84,85% dari 66 data citra uji.
2. Akurasi sistem dipengaruhi oleh kecocokan kombinasi sudut *GLCM* dengan *property GLCM* yang digunakan. Akurasi sistem dipengaruhi oleh *property GLCM* yang menghasilkan nilai fitur yang berdekatan dalam satu kelas, dan nilai yang cukup jauh antara satu kelas dengan kelas lain.
3. Akurasi sistem dipengaruhi oleh besar atau kecil nilai variabel *K*. Semakin besar nilai variabel *K*, akurasi sistem cenderung semakin berkurang. Akurasi sistem dipengaruhi oleh *distance K-NN* yang bisa menghitung jarak terdekat data uji dengan tiga kelas data latih dalam 1 dimensi data.

4. Dari analisis waktu komputasi dapat diketahui susunan parameter yang dapat memberikan performansi paling tinggi yaitu sudut $GLCM 90^0$, *property GLCM Difference Entropy*, nilai variabel $K 1$ dan *distance K-NN Euclidean* dengan waktu komputasi 0,63ms.

Daftar Pustaka:

- [1] Rangkuti, Nurul Hikmah (2007). *Perbedaan Leukoplakia dan Hairy Leukoplakia di Rongga Mulut*. Medan: Fakultas Kedokteran Gigi Universitas Sumatera Utara.
- [2] *www.doctorfungus.org* diakses pada tanggal 7 Juni 2015
- [3] Wijaya, Marvin Ch dan Agus Prijono. 2007. *Pengolahan Citra Digital Menggunakan Matlab, Image Processing Toolbox*. Bandung: Informatika.
- [4] Kadir, Abdul dan Adhi Susanto. 2013. *Teori dan Aplikasi Pengolahan Citra*. Yogyakarta: ANDI.
- [5] Saputra, Wahyu. 2015. *Peningkatan Kualitas Citra Periapikal Radiograf Pada Proses Deteksi Pulpitis*. Bandung: Fakultas Teknik Elektro Telkom University.
- [6] Putra, Toni Wijanarko Adi. 2013. *Pengenalan Wajah Dengan Matriks Kookurensi Aras Keabuan dan Jaringan Syaraf Tiruan Probabilistik*. Semarang: Program Pascasarjana Universitas Diponegoro.
- [7] Ekasari, Yunidha. 2015. *Diagnosis Kanker Serviks Menggunakan Model Recurrent Neural Network (RNN) Berbasis Graphical User Iinterface (GUI)*. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta.
- [8] Prasetyo, Eko. 2011. *Pengolahan Citra Digital dan Aplikasinya Menggunakan Matlab*. Yogyakarta: ANDI.

