

# IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK MENGENALI POLA CITRA DALAM MENDETEKSI PENYAKIT KULIT

**Sebastian Rori Listyanto**

Program Studi Teknik Informatika S1 Fakultas Ilmu Komputer  
Universitas Dian Nuswantoro Semarang  
Email: 111201005422@mhs.dinus.ac.id

Manusia memiliki kemampuan untuk mengklasifikasikan gambar ke dalam kelas yang sesuai. Kemampuan ini akan sangat berguna bila diterapkan dalam pada mesin seperti komputer. Tujuan dari penelitian ini adalah untuk mengimplementasikan *k-Nearest Neighbor* (k-NN) dan *Gray Level Coocurrence Matrix* (GLCM) pada Matlab agar bisa mengklasifikasikan gambar penyakit kulit ke dalam 3 kelas yang telah ditentukan yaitu sehat, panu dan skabies menggunakan analisis tekstur. Dataset berjumlah 70 citra untuk tiap kelas sehingga jumlah dari seluruh kelas adalah 210 citra. Citra yang berupa RGB 50 x 50 piksel diubah menjadi *grayscale* kemudian hasil data *grayscale* *discaling* menjadi 6 level dan dicari GLCM dengan derajat arah 0°, 45°, 90°, 135° dan jarak piksel 1, 2, 3, 4, 5 sesuai dengan uji coba. Kemudian diekstrak fitur-fitur yang terdapat dalam gambar. Fitur yang digunakan di penelitian ini adalah kontras, korelasi, energi dan homogenitas. Penelitian ini menggunakan 60 kali uji coba. Uji coba ini dilakukan dengan mencari akurasi klasifikasi dari setiap kemungkinan kombinasi sudut, jarak dan jumlah data *training* dan *testing* yang berbeda-beda. Kemudian dicari rata-rata dari semua uji coba berdasarkan sudut, jarak dan jumlah data *training* dan *testing* sehingga dapat diketahui pengaruh ketiganya terhadap akurasi klasifikasi. Pada hasil klasifikasi didapatkan akurasi tertinggi dengan menggunakan sudut 45°, jumlah data *training* dan *testing* 150 dan 60 serta jarak 3 piksel.

Kata kunci: **k-Nearest Neighbor, GLCM, Matlab, Klasifikasi**

## I. PENDAHULUAN

### A. Latar Belakang

Penglihatan manusia dapat melakukan hal-hal yang menakjubkan seperti mengenali orang/objek, navigasi dalam menghindari rintangan ataupun mengenali *mood* di dalam sebuah adegan. Lain halnya dengan mesin yang dalam konteks ini adalah komputer. Untuk meniru penglihatan seperti manusia, komputer memerlukan sensor yang berfungsi layaknya mata pada manusia dan program komputer yang berfungsi sebagai pemroses data dari sensor. *Computer vision* merupakan ilmu yang menggunakan *image processing* untuk membuat keputusan berdasarkan citra yang didapat dari sensor [1], [2], [3]. Dengan kata lain *computer vision* bertujuan untuk membangun sebuah mesin pandai yang dapat melihat.

Kulit merupakan bagian tubuh yang paling utama yang perlu diperhatikan. Kulit juga merupakan organ tubuh paling besar yang melapisi seluruh bagian tubuh, membungkus daging dan organ-organ yang ada di dalamnya. Kulit memiliki fungsi melindungi bagian tubuh dari berbagai macam gangguan dan rangsangan dari luar. Mengingat pentingnya kulit sebagai pelindung organ-organ tubuh di dalamnya, maka kulit perlu dijaga kesehatannya.

Analisis tekstur merupakan salah satu teknik analisis citra berdasarkan anggapan bahwa citra dibentuk oleh variasi intensitas piksel, baik citra keabuan (*grayscale*) maupun citra warna [4]. Variasi intensitas piksel pada sebuah bidang citra ini yang disebut tekstur. Tekstur merupakan salah satu elemen dasar citra yang berupa ciri-ciri atau sifat-sifat yang terdapat dalam citra. Oleh karena itu, tekstur

dapat digunakan untuk klasifikasi terhadap citra yaitu untuk mengidentifikasi dan mengenali suatu citra dengan tekstur tertentu berdasarkan satu set kelas atau klasifikasi tekstur yang diberikan. Setiap daerah pada citra tekstur tersebut memiliki karakteristik yang unik. Untuk mendapatkan karakteristik tersebut diperlukan ekstraksi ciri yang berguna untuk mengambil ciri-ciri yang berbeda dari tiap daerah pada citra tekstur tersebut.

Fitur-fitur yang dihasilkan oleh metode ekstraksi ciri kemudian akan digunakan sebagai masukan proses klasifikasi. Metode klasifikasi yang digunakan dalam penelitian ini adalah *k-Nearest Neighbor* (k-NN). K-NN adalah sebuah metode klasifikasi yang tangguh terhadap data *training* yang noisy dan efektif apabila datanya besar. Algoritma k-NN memiliki konsistensi yang kuat dan ketika jumlah data mendekati tak hingga, algoritma ini menjamin *error rate* yang tidak lebih dari dua kali *Bayes error rate* (*error rate* minimum untuk distribusi data tertentu) [5].

### B. Tujuan

Tujuan penelitian ini adalah untuk mengimplementasikan *k-Nearest Neighbor* dalam klasifikasi citra penyakit kulit.

### C. Batasan Masalah

1. Citra penyakit kulit yang digunakan memiliki ukuran 50 x 50 piksel.
2. Dataset yang digunakan adalah data citra yang terdapat pada website dermnet.
3. Penyakit yang diteliti adalah panu dan skabies.
4. Citra diklasifikasikan kedalam 3 kelas yaitu sehat, panu dan skabies.
5. Jarak piksel yang digunakan adalah 1, 2, 3, 4, 5 piksel.
6. Arah sudut yang digunakan adalah 0°, 45°, 90°, 135°.
7. Skala keabuan yang digunakan dalam mencari matriks GLCM adalah 6.
8. Jumlah dataset yang digunakan adalah 70 citra untuk tiap kelasnya sehingga

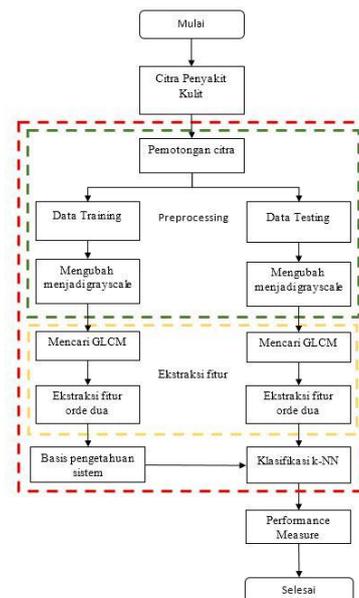
jumlah seluruh dataset untuk 3 kelas adalah 210 citra.

9. Input *k-Nearest Neighbor* adalah hasil ekstraksi 4 fitur dari citra kulit yaitu kontras, korelasi, energi dan homogenitas.
10. Jumlah tetangga terdekat yang digunakan dalam klasifikasi k-NN adalah 3.

## II. METODE YANG DIUSULKAN

Langkah-langkah yang dikembangkan dalam penelitian ini adalah sebagai berikut.

1. Memilih citra yang akan diteliti dari dermnet dan mengunduhnya.
2. Melakukan pemotongan bagian citra yang akan diteliti menjadi ukuran 50 x 50 piksel.
3. Memisahkan citra *training* dan *testing* serta kelas-kelasnya.
4. Mengubah citra menjadi grayscale.
5. Mencari *Gray Level Coocurrence Matrix* (GLCM) dari citra.
6. Mengekstraksi fitur dari GLCM citra
7. Melakukan *training* citra menggunakan *k-Nearest Neighbor*
8. Menghitung akurasi menggunakan *recognition rate*.



**Gambar 3.1 Diagram proses klasifikasi penyakit kulit**

Pada gambar dapat dilihat untuk perlakuan data *testing* dan data *training* adalah sama. Data *training* digunakan untuk memberi pengetahuan pada sistem, sehingga sistem dapat memiliki pengetahuan mengenai macam-macam penyakit kulit yang dipilih. Data *testing* digunakan untuk menguji sistem dan algoritma yang digunakan.

Pertama kali data *training* akan diproses, setelah selesai kemudian sistem akan memproses data *testing*. Data *training* dan data *testing* akan diolah dengan cara yang sama. Masing-masing data *training* dan data *testing* akan melalui tahap *preprocessing* dan ekstraksi fitur. Setelah tahap ekstraksi fitur, data *training* dan data *testing* akan diklasifikasikan. Pada tahap akhir, setelah data terklasifikasi atau data *training* dikenali polanya, sistem akan mengukur akurasi dari algoritma yang digunakan.

### III. IMPLEMENTASI

#### A. Prapengolahan Citra

Langkah prapengolahan untuk data *testing* dan data *training* adalah sama. Adapun langkah-langkah dalam prapengolahan adalah sebagai berikut.

1. Menyiapkan citra penyakit kulit  
Siapkan citra penyakit kulit panu dan skabies yang telah di download dari dermnet.
2. Menyamakan dimensi citra  
Citra yang telah didapatkan dari dermnet masih berupa citra berdimensi besar yaitu 720 x 480 dan informasi yang diberikan masih terlalu umum, sehingga diperlukan pemotongan sesuai bagian-bagian yang akan diteliti saja. Dalam penelitian ini citra dipotong menjadi ukuran 50 x 50 piksel. Hal ini dimaksudkan agar beban sistem dalam melakukan pemrosesan data menjadi lebih ringan.
3. Membagi menjadi data *training* dan *testing*

Setiap kelas citra yang telah dipotong-potong akan dibagi menjadi data *training* dan data *testing*. Data akan diambil secara acak dari tiap kelas dan dikumpulkan dalam folder *training* dan *testing*. Masing-masing folder terdiri dari 3 kelas yaitu citra kulit sehat, kulit panu dan kulit skabies. Masing-masing citra diberi nama dengan penomoran (1.jpg, 2.jpg) agar proses dalam program dapat dilakukan dengan perulangan.

#### 4. Mengubah menjadi *grayscale*

Citra yang diperoleh dari dermnet masih berupa citra RGB yang terdiri dari 3 warna (merah, hijau dan biru) untuk setiap pikselnya. Perubahan citra menjadi *grayscale* bertujuan mengubah 3 nilai warna (merah, hijau dan biru) menjadi hanya 1 nilai (*grayscale*) sehingga proses ekstraksi fitur dari GLCM bisa didapatkan.

#### B. Ekstraksi Fitur

Pada tahap ekstraksi fitur tidak ada perbedaan perlakuan antara data *training* dan data *testing*. Ekstraksi fitur dalam penelitian ini menggunakan 2 tahap yaitu:

1. Pembentukan matriks GLCM  
Setelah mendapatkan nilai abu-abu pada *preprocessing*, kemudian dilakukan proses *scaling* ke dalam 6 skala. Setelah didapatkan hasil *scaling*, kemudian dihitung matriks GLCM untuk sudut yang diuji coba yaitu 0°, 45°, 90° atau 135°. Dari perhitungan ini akan didapatkan matriks GLCM 6 x 6.
2. Ekstraksi Fitur Orde dua  
Setelah mendapatkan matriks GLCM, kemudian dicari normalisasinya sehingga jumlah dari semua elemennya adalah satu. Setelah itu dicari fitur dari matriks GLCM.

#### C. Basis Pengetahuan

1. Representasikan masing-masing hasil ekstraksi fitur menjadi matriks berdimensi satu.

2. Kelompokkan data citra *training* menjadi satu matriks dan data citra *testing* menjadi 1 matriks. Masing-masing baris pada matriks merepresentasikan satu citra penyakit kulit.
3. Sebelum matriks tersebut digunakan sebagai basis pengetahuan, terlebih dahulu buat grup untuk menandakan citra-citra yang terdapat di dalam data *training* sesuai dengan masing-masing jenis penyakit. Dalam penelitian ini dibuat kelompok sebagai berikut:

- 100 = Kulit Sehat  
 200 = Kulit Panu  
 300 = Kulit Skabies

#### D. Klasifikasi Menggunakan *k*-Nearest

##### *Neighbor*

Data *training* dan data *testing* yang sudah dijadikan dalam satu matriks akan dibandingkan menggunakan metode *k-Nearest Neighbor* dengan pencarian jarak *Euclidean*. Pada penelitian ini dilakukan 60 kali uji coba dengan kombinasi jarak, sudut dan jumlah data *training* dan data *testing* yang berbeda-beda.

##### E. Penghitungan Rata Akurasi

Perhitungan rata akurasi dilakukan untuk melihat seberapa tepat algoritma ekstraksi fitur orde dua dengan GLCM dan sejauh mana jarak piksel dan arah sudut mempengaruhi akurasi. Perhitungan rata akurasi menggunakan algoritma *recognition rate*. Peneliti melakukan tiga kali perhitungan rata akurasi yaitu berdasarkan sudut, jumlah data *training* dan jarak antar piksel.

## IV. HASIL DAN PEMBAHASAN

#### A. Klasifikasi Menggunakan *k*-Nearest

##### *Neighbor*

**Tabel 4.1 Hasil uji coba (data *training*= 90, data *testing*= 120)**

	1	2	3	4	5
0	50,83	42,50	46,66	49,16	47,50
45	54,16	50	47,5	47,5	47,5
90	44,16	43,33	42,5	50,83	45
135	39,16	45	43,33	50	52

Dari tabel diatas akurasi tertinggi didapat jika menggunakan jarak 1 piksel dan sudut 45° dengan akurasi sebesar 54,16% dan akurasi terendah didapat pada uji coba dengan menggunakan jarak 1 piksel dan sudut 135° dengan akurasi sebesar 39,16%.

**Tabel 4.2 Hasil uji coba (data *training*= 120, data *testing*= 90)**

	1	2	3	4	5
0	51,11	45,55	53,33	52,22	54,44
45	56,66	58,88	62,22	46,66	40
90	43,33	45,55	54,44	56,66	53,33
135	45,55	55,55	47,77	50	48,88

Dari tabel diatas akurasi tertinggi didapat jika menggunakan jarak 3 piksel dan sudut 45° dengan akurasi sebesar 62,22% dan akurasi terendah didapat pada uji coba dengan menggunakan jarak 5 piksel dan sudut 45° dengan akurasi sebesar 40%.

**Tabel 4.3 Hasil uji coba (data *training*= 150, data *testing*= 60)**

	1	2	3	4	5
0	51,66	43,33	65	56,66	58,33
45	58,33	61,66	60	50	50
90	48,33	53,33	55	60	48,33
135	43,33	56,66	48,33	45	51,66

Dari tabel diatas akurasi tertinggi didapat jika menggunakan jarak 3 piksel dan sudut 0°

dengan akurasi sebesar 65% dan akurasi terendah didapat pada 2 uji coba yang menggunakan jarak 1 piksel dan sudut 135° serta jarak 2 piksel dan sudut 0° dengan akurasi sebesar 43,33%.

### B. Penghitungan Rata Akurasi

Perhitungan rata akurasi dilakukan untuk melihat seberapa tepat algoritma ekstraksi fitur orde dua dengan GLCM dan sejauh mana jarak piksel dan arah sudut mempengaruhi akurasi. Perhitungan rata akurasi menggunakan algoritma *recognition rate*. Peneliti melakukan tiga kali perhitungan rata akurasi yaitu berdasarkan sudut, jumlah data *training* dan jarak antar piksel. Hasil perhitungan rata akurasi untuk setiap sudut bisa dilihat pada tabel 4.4 berikut.

**Tabel 4.4 Rata-rata akurasi berdasarkan sudut**

Sudut	Rata-Rata Akurasi
0°	51,22
45°	52,74
90°	49,61
135°	48,18

Rata akurasi berdasarkan sudut didapatkan dengan menghitung nilai rata-rata dari setiap 60 uji coba yang memiliki sudut yang sama. Dari tabel diatas didapatkan akurasi tertinggi dengan menggunakan sudut 45° dan akurasi terendah dengan sudut 135°. Sedangkan untuk rata akurasi berdasarkan jumlah data *training* bisa dilihat pada tabel 4.5 berikut.

**Tabel 4.5 Rata-rata akurasi berdasarkan jumlah data *training-testing***

Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>	Rata-Rata Akurasi
90	120	46,96
120	90	51,11
150	60	53,25

Rata akurasi berdasarkan jumlah data *training* dan *testing* didapatkan dengan menghitung rata-rata akurasi dari uji coba yang memiliki jumlah data *training* dan *testing* yang sama. Berdasarkan tabel di atas, rata-rata nilai akurasi meningkat sesuai dengan peningkatan jumlah data *training*. Akurasi tertinggi ditunjukkan oleh uji coba yang menggunakan jumlah data *training* 150 dan jumlah data *testing* 60.

Selanjutnya untuk rata akurasi berdasarkan jarak bisa terlihat pada tabel 4.6 berikut.

**Tabel 4.6 Rata-rata akurasi berdasarkan jarak piksel**

Jarak	Rata-Rata Akurasi
1	48,88
2	50,11
3	52,17
4	51,22
5	49,79

Rata akurasi berdasarkan jarak didapatkan dengan menghitung nilai rata-rata akurasi berdasarkan uji coba yang memiliki jarak piksel yang sama. Berdasarkan tabel di atas, rata-rata nilai akurasi yang tertinggi didapatkan dengan menggunakan jarak 3 piksel.

## V. PENUTUP

### A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Berdasarkan uji coba yang telah dilakukan menggunakan sudut, jarak serta jumlah data *training* dan *testing* yang berbeda-beda maka tingkat pengenalan citra penyakit tertinggi ditunjukkan dengan menggunakan sudut 45°, jumlah data *training* dan *testing* 150 dan 60 serta jarak 3 piksel.
2. Rata-rata akurasi meningkat jika menggunakan kombinasi jumlah data

*training* meningkat dan jumlah data *testing*nya menurun.

3. Rata-rata akurasi tertinggi didapat dengan menggunakan sudut  $45^\circ$  dan terendah dengan menggunakan sudut  $135^\circ$ .
4. Rata-rata akurasi tertinggi didapat dengan menggunakan jarak 3 piksel dan akurasi terendah dengan jarak 1 piksel.
5. Rata-rata akurasi tertinggi didapat dengan menggunakan kombinasi data *training* dan *testing* 150 dan 60 dan akurasi terendah dengan kombinasi data *training* dan *testing* 90 dan 120.

#### B. Saran

Saran peneliti untuk mengembangkan tugas akhir ini antara lain sebagai berikut:

1. Penelitian dapat dilanjutkan dengan memperbanyak jumlah citra *training* agar akurasi klasifikasi bisa lebih akurat.
2. Penelitian selanjutnya diharapkan dapat mengklasifikasikan jenis penyakit (kelas) yang lebih beragam.
3. Penelitian dapat dikembangkan dengan mengubah algoritma yang diterapkan dalam ekstraksi fitur atau tahap klasifikasi untuk kemudian dibandingkan dengan algoritma yang menerapkan GLCM dan ekstraksi fitur orde dua dalam penelitian ini.

#### DAFTAR PUSTAKA

- [1] D. Forsyth dan J. Ponce, *Computer Vision: A modern Approach* 2nd ed, Pearson Education, Ltd, 2011.
- [2] L. Shapiro dan G. Stockman, *Computer Vision*, Prentice Hall, 2001.
- [3] R. Szeliski, *Computer Vision: Algorithms and Application*, Springer, 2011.
- [4] H. Wibawanto, "Analisis Tekstur untuk Diskriminasi Massa Kistik dan Non Kistik pada Citra Ultrasonografi," Skripsi Program Pasca Sarjana Fakultas Teknik UGM, Yogyakarta, 2011.
- [5] A. Sukma, D. Ramadhan, B. P. Santoso, T. R. Sari dan N. M. A. K. Wiraswari, "k-Nearest Neighbor Information Retrieval (Sistem Temu Kembali Informasi)," Universitas Airlangga Fakultas Sains dan Teknologi, Surabaya, 2014.
- [6] A. J. Arriawati, I. Santoso dan Y. Christyono, "Klasifikasi Citra Tekstur Menggunakan k-Nearest Neighbor Berdasarkan Ekstraksi Ciri Metode Matriks Kookurensi," Skripsi Teknik Elektro Universitas Diponegoro, Semarang, 2011.
- [7] Z. Budiarto, "Identifikasi Macan Tutul dengan Metode Grey Level Coocurent Matrix (GLCM)," *Jurnal Dinamika Informatika Universitas Stikubank*, Semarang, 2010.
- [8] Y. G. K, I. Santoso dan R. R. Isnanto, "Klasifikasi Citra dengan Matriks Ko-Okurensi Aras Keabuan (Gray Level Co-occurrenceMatrix-GLCM) Pada Lima Kelas Biji-Bijian," Skripsi Teknik Elektro Universitas Diponegoro, Semarang, 2011.
- [9] R. Munir, *Pengolahan Citra Digital dengan Pendekatan Algoritmik*, Bandung: Informatika Bandung, 2004.
- [10] D. Putra, *Sistem Biometrika Konsep Dasar, Teknik Analisis Citra dan Tahapan Membangun Aplikasi Sistem Biometrika*, Yogyakarta: Andi, 2009.
- [11] J. Hopewell, "The Skin: Its structure and Response to Ionizing Radiation," *Int. J. Radiat. Biol.*, no. 57, pp. 751-773, 1990.
- [12] B. Schottellius dan D. Schottellius, *Textbook of Physiology*, Saint Louis: The C.V. Mosby Company, 1973.
- [13] A. Kusuma, R. Isnanto dan I. Santoso, "Pengenalan Iris Mata Menggunakan

Pencirian Matriks Ko-Okurensi Aras Keabuan,” Skripsi Teknik Elektro Universitas Diponegoro, Semarang, 2011.

<http://www.dermnet.com>. [Diakses 16 March 2015].

- [14] F. Albregesten, “Statistical Texture Measures Computed from Gray Level Coocurrence Matrices,” Department of Informatics University of Oslo, Oslo, 2008.
- [15] Y. Febriyanto, “Pengklasifikasian Kualitas Keramik Berdasarkan Ekstraksi Fitur Tekstur Statistik,” Skripsi Teknik Informatika Universitas Gunadarma, Depok, 2012.
- [16] T. W. A. Putra, “Pengenalan Wajah dengan Matriks Kookurensi Aras Keabuan dan Jaringan Syaraf Tiruan Probabilistik,” Tesis Sistem Informasi Universitas Diponegoro, Semarang, 2013.
- [17] A. Kadir, N. L. E., A. Susanto dan P. Santosa, “Neural Network Application on Foliage Plant Identification,” *International Journal of Computer Application (0975-8887)*, vol. 29, no. 9, pp. 15-22, 2011.
- [18] F. M. Hana, “Sistem Identifikasi Biometrik Finger Knuckle Print Menggunakan Histogram Equalization dan Principal Component Analysis (PCA),” Skripsi Teknik Informatika Universitas Dian Nuswantoro, Semarang, 2014.
- [19] S. B. Palilati, “Hubungan Antara Higiene Pribadi dan Penggunaan Alat Pelindung Diri dengan Kejadian Penyakit Kulit Pada Pekerja Pengangkut Sampah Di Kota Gorontalo,” Skripsi Kesehatan Masyarakat Universitas Negeri Gorontalo, Gorontalo, 2012.
- [20] M. Thomas Habif, “Dermnet Skin Diseases Atlas,” Dermnet, 1998. [Online]. Available: