

Deteksi Otomatis Penyakit Kulit Menggunakan Algoritma *Naive Bayes*

Izzati Saila Hafsa¹, Pulung Nurtantio Andono²

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Jl. Nakula No. 5-11 Semarang-50131

E-mail : izzatisaila@gmail.com¹, maspapu@gmail.com²

Abstrak

Kulit merupakan organ tubuh terluar yang membatasi organ tubuh lain dengan lingkungan hidup manusia. Permukaan kulit manusia mengandung banyak bahan makanan (nutrisi) untuk pertumbuhan organisme, antara lain lemak, bahan-bahan yang mengandung nitrogen, dan mineral, sehingga kulit manusia mudah terinfeksi bakteri, jamur ataupun virus yang menyebabkan berbagai macam penyakit kulit. Dalam ilmu medis, salah satu cara untuk mendiagnosa seseorang berpenyakit atau tidak yaitu dengan melihat secara langsung kulit orang yang bersangkutan. Penelitian ini mengimplementasikan metode klasifikasi dan *image processing* untuk menghasilkan perangkat lunak yang dapat mendiagnosa citra kulit berpenyakit dan mengklasifikasikan tipe penyakit tersebut serta menganalisa performansi metode *naive bayes* dan ekstraksi fitur tekstur berbasis histogram yang diterapkan. Berdasarkan hasil penelitian ditunjukkan bahwa, klasifikasi jenis penyakit kulit (Acne, Herpes, Eczema) dengan menggunakan metode *naive bayes* berdasarkan tekstur pada citra dapat digunakan sebagai penentu penyakit kulit secara visual. Hasil penelitian menunjukkan bahwa klasifikasi jenis penyakit kulit dengan menggunakan metode *naive bayes* berdasarkan tekstur pada citra memperoleh nilai akurasi sebesar 73,33%.

Kata kunci : *Naive Bayes, Image Processing, Histogram, Penyakit Kulit*

Abstract

The outer skin is an organ that restrict other organs in the human environment. The surface of human skin contains fat food ingredients (nutrients) to pertubuhan organisms, such as fat, substances that contain nitrogen, and minerals, so that the human skin is infected with bacteria, fungi and viruses that cause various skin diseases. In medical science, one way to remedy diagnose diseased or not a person is to see the person's skin lngsung innovative. This study implements classification method and image processing software to produce images that can diagnose diseased skin and classify the type of the disease seta analyze performance and feature extraction methods *naivebayes* histogram-based texture is applied. Based on the results of the study indicated that, the classification of types of skin diseases (acne, herpes, eczema) using *Naive Bayes* method based on image texture can be used as a determinant of skin diseases visually. 1 The results showed that the classification of types of skin diseases using *Naive Bayes* method based on image texture scored an accuracy of 73.33%.

Keyword : *Naive Bayes, Image Processing, Histogram, Skin Disease*

1. Pendahuluan

Kulit adalah organ terbesar dalam tubuh manusia, yang membantu untuk menutupi otot, tulang dan semua bagian tubuh. Fungsi kulit di tubuh manusia memiliki kepentingan yang lebih besar karena perubahan kecil dalam fungsi tersebut dapat mempengaruhi bagian lain dari tubuh. kulit terkena lingkungan luar sehingga penyakit dan infeksi terjadi lebih

pada kulit. Deteksi dini penyakit kulit lebih kompleks dengan dokter kulit berpengalaman. Dengan menggabungkan pengolahan citra digital untuk deteksi kanker kulit, memungkinkan untuk melakukan diagnosis tanpa kontak fisik dengan kulit. Untuk alasan ini, pengembangan *Computer-Aided Diagnosis System (CADs)* telah menjadi area utama penelitian di

bidang medis. *Machine learning* memainkan peran penting dalam bidang medis untuk otomatisasi banyak proses. Telah menunjukkan bahwa dermoscopy sebenarnya menurunkan akurasi diagnostik di tangan dokter kulit berpengalaman.[1]

Dalam melakukan proses klasifikasi otomatis maka sebelumnya perlu dilakukan proses analisis tekstur untuk mengidentifikasi parameter penentu dari penyakit kulit. Analisis tekstur pada umumnya membutuhkan tahapan fitur tekstur, dimana metode ini terbagi dalam 3 golongan: metode statis, metode struktural dan metode spektral [2]. Dalam penelitian ini metode statis dipilih dalam tahapan analisis tekstur. Metode ini menggunakan perhitungan statistika untuk membentuk fitur.

Metode *Histogram* merupakan metode statis orde satu untuk memperoleh fitur tekstur. Fitur-fitur yang dapat dikenali melalui metode *histogram* di antaranya adalah rerata intensitas, deviasi standar, *skewness*, energi, entropi dan kehalusan[2]. *Naive Bayes Clasifier* (NBC) merupakan salah satu metode pembelajaran *supervised* yang mudah, efisien, efektif dan handal menangani derau data seperti atribut yang tidak relevan.

2. Landasan Teori

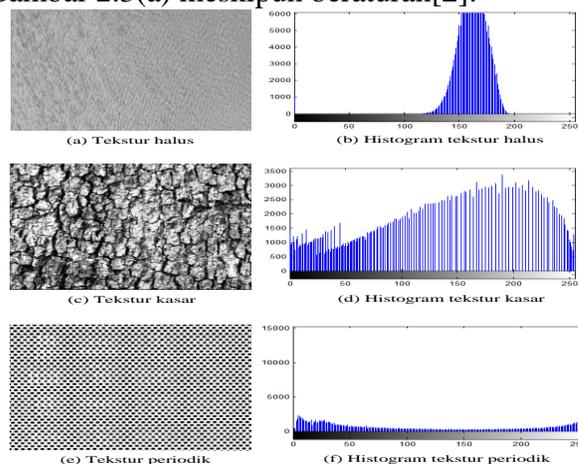
2.1 Pengertian Dasar Citra

Citra (*image*) adalah gambar pada bidang dwimatra (dua dimensi). Ditinjau dari sudut pandang matematis, citra merupakan fungsi menerus (*continue*) dari intensitas cahaya pada bidang dwimatra. Sumber cahaya menerangi objek, objek memantulkan kembali sebagian dari berkas cahaya tersebut. Pantulan cahaya ini ditangkap oleh alat optik, misalnya mata pada manusia, kamera dan pemindai (*scanner*), sehingga bayangan objek yang disebut citra tersebut terekam[10]

Citra sebagai keluaran suatu sistem perekaman data dapat bersifat optik berupa foto, bersifat analog berupa sinyal – sinyal video seperti gambar pada monitor televisi, atau bersifat digital yang dapat langsung disimpan pada suatu media penyimpan. Suatu citra analog tidak dapat secara langsung diproses oleh komputer. Untuk memproses citra analog menggunakan komputer maka sebelumnya citra analog harus diubah menjadi citra digital

2.2 Ekstraksi Fitur Tekstur Berbasis Histogram

Metode *histogram* merupakan metode statis orde satu untuk memperoleh fitur tekstur. fitur-fitur yang dapat dikenali melalui metode *histogram* di antaranya adalah rerata intensitas, deviasi standar, *skewness*, energi, entropi dan kehalusan. *Histogram* dari 3 buah citra yang mengandung tekstur yang berbeda, terdapat di Gambar 2.3 Gambar 2.3(a) menunjukkan bahwa citra dengan tektur halus memiliki daerah perubahan intensitas yang sempit. Sebaliknya, citra yang kasar memiliki kontras yang tinggi, ditandai dengan jangkauan intensitas yang lebar seperti pada Gambar 2.3(c) dan (d). Menurut penglihatan, citra dalam Gambar 2.3(e) juga termasuk kasar dibandingkan dengan citra pada Gambar 2.3(a) meskipun beraturan[2].



2.3 Algoritma Naive Bayes

Naive Bayes Classifier (NBC) disebut juga *Bayesian Classification* merupakan metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan dari suatu class. *Naive bayes classifier* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa *Decision Tree* dan *Neural Network*.

Secara umum cara kerja NBC berdasarkan bentuk umum teorema bayes yaitu [4]:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (7)$$

Dimana dalam hal ini:

C_i = hipotesis data X merupakan suatu kategori/class spesifik.

X = data dengan class yang belum diketahui.

$P(C_i|X)$ = probabilitas hipotesis C_i berdasarkan kondisi X (*posteriori probability*).

$P(X|C_i)$ = probabilitas X berdasarkan kondisi pada hipotesis C_i .

$P(C_i)$ = probabilitas hipotesis C_i (*prior probability*).

$P(X)$ = probabilitas dari data X .

Untuk data bernilai nominal maka penggunaan metode sebelumnya akan berjalan langsung sebagai mana mestinya. Namun untuk data numerik maka kita perlu sedikit lebih pekerjaan, nilai numerik biasanya ditangani dengan asumsi bahwa mereka memiliki probabilitas distrbusi “normal” atau “*Gaussian*” dengan menghitung nilai dari rerata μ dan standar deviasi σ pada setiap class. Didefinisikan dengan [17][18]:

$$g(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2(\sigma^2)}} \quad (8)$$

sehingga dapat disimpulkan pencarian probabilitas X berdasarkan kondisi pada hipotesis C_i :

$$P(X_k|C_i) = g(x_k, \mu_i, \sigma_i) \quad (9)$$

Pengukuran bertujuan untuk mengetahui kinerja dari “Algoritma *Naive Bayes* dalam menentukan jenis penyakit kulit” dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Untuk mengetahui kinerjanya maka diperlukan suatu metode perhitungan kinerja yaitu menggunakan akurasi.

Perhitungan nilai akurasi, dapat didefinisikan dengan persamaan:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Dalam hal ini dapat dijelaskan bahwa :

TP = jumlah dari data yang terklasifikasi di kelas yang benar.

TN = jumlah dari data yang terklasifikasi di kelas sebaliknya yang benar.

FP = jumlah data yang dianggap berada di kelas yang benar oleh sistem padahal seharusnya data tersebut berada di kelas yang sebaliknya.

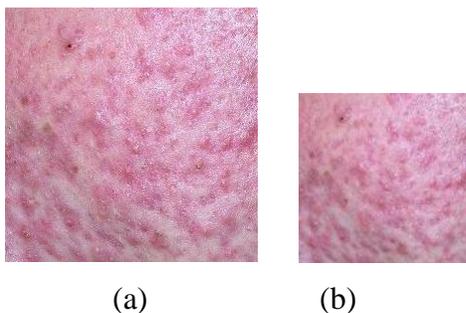
FN = jumlah data yang dianggap berada di kelas yang sebaliknya oleh sistem padahal seharusnya data tersebut berada di kelas yang benar.

3 Pembahasan

Dalam melakukan eksperimen penelitian ini diperlukan sejumlah 105 citra penyakit kulit yang digunakan sebagai dataset. Sebelum dilakukan proses klasifikasi penyakit kulit, maka terlebih dahulu dilakukan proses penyortiran citra sesuai jenis penyakit dan sesuai kelompok datanya.

3. Normalisasi Citra Penyakit Kulit

Citra penyakit kulit memiliki dimensi yang besar, sehingga perlu dilakukan normalisasi kedalam ukuran tertentu untuk semua data. Hal ini bertujuan untuk menyamakan ukuran dan menyamakan perhitungan serta meminimalkan konsumsi memori. Ukuran yang ditentukan untuk penelitian ini adalah 128x128 piksel.



3.2 Klasifikasi Citra Penyakit Kulit

Pada penelitian ini proses eksperimen klasifikasi jenis penyakit kulit menggunakan metode *naive bayes* dilakukan dengan cara membuat sebuah *prototype* sistem menggunakan Bahasa pemrograman PHP. Berdasarkan persamaan (7) yang digunakan dalam melakukan perhitungan *naive bayes* maka mula-mula akan dilakukan perhitungan terhadap *prior probability* $P(C_i)$ setiap *class*. $P(C_i)$ dalam penelitian ini merupakan nilai probabilitas dari jumlah data latih yang terdiri dari 3 *class* yang berisikan 30 data dengan seluruhan data tiap *class*. Setelah melakukan perhitungan maka didapatkan nilai dari $P(C_i)$ setiap *class* sebagai berikut :

$$P(\text{penyakit} = \text{"Acne"}) = 30 / 90 = 0,3333$$

$$P(\text{penyakit} = \text{"Herpes"}) = 30 / 90 = 0,3333$$

$$P(\text{penyakit} = \text{"Eczema"}) = 30 / 90 = 0,3333$$

Setelah selesai melakukan perhitungan $P(C_i)$ maka selanjutnya akan dilakukan

perhitungan $P(X_k|C_i)$ setiap atribut berdasarkan *class*. Pada subbab sebelumnya telah dijelaskan untuk data bertipe numerik maka *naive bayes* akan melakukan perhitungan fungsi distribusi normal atau *gaussian* melalui persamaan (8) untuk mendapatkan $P(X_k|C_i)$. persamaan (8) maka sebelumnya perlu dilakukan perhitungan nilai dari rerata μ dan standar deviasi σ pada setiap *class*. Untuk menghitung *mean* (μ) dan *standar deviasi* (σ) maka digunakan persamaan berikut :

$$\mu = \frac{\sum_{i=1}^n X}{n} \quad (11)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X - \mu)^2}{n - 1}}$$

Dimana X = nilai data dari atribut yang digunakan dan n = nilai dari total data setiap *class*

Untuk melakukan pembuktian kebenaran terhadap hasil yang diberikan sistem maka penulis juga melakukan contoh perhitungan $P(X_k|C_i)$ secara manual sebagai berikut :

$$P(\text{mean} = 151,1646 | \text{Acne}) = \frac{1}{20,67\sqrt{2(3.14)}} 2,718282^{-\frac{(151,1646 - 150,584)^2}{2(20,67)^2}}$$

$$P(\text{mean} = 151,1646 | \text{Acne}) = 0,01929$$

Dengan cara mengalikan $P(X_k|C_i)$ dan $P(C_i)$ setiap *class* maka nilai *posterior probability* $P(C_i|X)$ setiap *class* sudah dapat diketahui.

3.3 Evaluasi Dan Validasi

Dari eksperimen kepada 15 data citra uji yang telah dilakukan hasilnya kemudian dimasukkan ke dalam tabel *confusion matrix* untuk memudahkan dalam menghitung nilai akurasi sistem klasifikasi penyakit kulit

Tabel 4.1 *Confusion Matrix* untuk metode yang diusulkan

	True Acne	True Herpes	True Eczema
Pred Acne	4	1	0
Pred Herpes	0	3	1
Pred Eczema	1	1	4

$$\begin{aligned} \text{Akurasi} &= \frac{11}{15} = 0,7333 \times 100\% \\ &= 73,33\% \end{aligned}$$

Berdasarkan hasil perhitungan diatas didapatkan akurasi sistem dalam melakukan klasifikasi mencapai 73,33%. Hal tersebut dapat disimpulkan bahwa penggunaan metode *naive bayes* dalam menentukan penyakit kulit berdasarkan ekstraksi fitur *statistical histogram* dapat digunakan sebagai sistem klasifikasi penyakit kulit. Namun hasil akurasi yang diberikan kurang maksimal jika dibandingkan dengan penelitian [5]. Hal tersebut dapat terjadi karena faktor kurangnya data penelitian sehingga berpengaruh terhadap hasil klasifikasi.

4 Penutup

Berdasarkan penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa klasifikasi jenis penyakit kulit dengan menggunakan metode *naive bayes* berdasarkan tekstur pada citra memberikan nilai akurasi yang berada dibawah penelitian sebelumnya dan memperoleh nilai akurasi sebesar 73,33%.

DaftarPustaka

- [1] S. R, M. Suhil, and D. S. Guru, "Segmentation and Classifications of Skin Lesions for Disease Diagnosis," in *International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)*, Mysore, 2015, pp. 76-85.
- [2] A. Kadir and A. Susanto, *Pengolahan Citra Teori dan Aplikasi*. Yogyakarta, Indonesia: Andi, 2013.
- [3] Y. Yang and G. I. Webb, "On Why Discretization Works for Naive-Bayes Classifier," 2003.
- [4] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta, Indonesia: ANDI, 2006.
- [5] A. K. Mitra and R. Parekh, "Automated Detection of Skin Diseases Using Texture Features," *International Journal of Engginering Science and Technology (IJEST)*, vol. 3, pp. 4801-4808, Jun. 2011.
- [6] S. Sergyan, "Color Histogram Features Based Image Classification In Content-Based Image Retrieval Systems," *International Symposium on Applied Machine Intelligence and Informatics*, 2008.
- [7] A. Kadir, "A Model of Plant Identification System Using GLCM, Lacunarity and Shen Features," 2014.
- [8] E. R. Anandita, "Klasifikasi Tebu dengan Menggunakan Algoritma Naive Bayes Classification pada Dinas Kehutanan Dan Perkebunan Pati," 2014.
- [9] Y. Unknown, A. Mumpuni, and W. Unknown, *Cara Jitu Mengatasi Jerawat*. Jogjakarta, Indonesia: Andi, 2010.
- [10] R. Munir, *Pengolahan Citra Digital dengan Pendekatan Algoritmik*. Jakarta, Indonesia: Informatika, 2004.
- [11] T. Sutojo, E. Mulyanto, V. Suhartono, O. D. Nurhayati, and Wijanarto, *Teori*

- Pengolahan Citra Digital*. Yogyakarta, Indonesia: ANDI, 2009.
- [12] O. Marques and B. Furht, *Content Based Image and Video Retrieval*. Florida, United States: Kluwer Academic, 2002.
- [13] A. D. Kulkarni, *Artificial Neural Networks for Image Understanding*. Newyork, Unites States of America: Van Nostrand Reinhold, 1994.
- [14] T. Archary and A. K. Ray, *Image Processing Principles and Applications*. New Jersey, United States of America: John Wiley & Sons, Inc., 2005.
- [15] M. Petrou and P. G. Sevilla, *Image Processing Dealing With Texture*. Chichester: John Wiley & Sons Inc, 2006.
- [16] E. Prasetyo, *DATA MINING - Mengolah Data Menjadi Informasi Menggunakan Matlab*, A. Sahala, Ed. Yogyakarta, Indonesia: ANDI, 2014.
- [17] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, United States of America: Morgan Kaufmann, 2011.
- [18] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. Waltham, United States of America: Morgan Kaufmann, 2012.
- [1] Pratiwi, R. R., Fillah Fithra Dieny. 2014. Hubungan Skor Frekuensi Diet Bebas Casein Dengan Skor Perilaku Autis. *Journal of Nutrition College*. Vol. 3., No. 1. Hal : 34-42.
- [2] Nurmansyah, W., Sri Hartati. 2013. Prototipe Sistem Pakar Penentu Jenis Gangguan Psikologi Klinis Menggunakan *Forward Chaining* dan *Formula Bayes*. Yogyakarta. FMIPA UGM.
- [3] Rohman, F. F., Ami Fauziah. 2008. Rancang Bangun Aplikasi Sistem Pakar Untuk Menentukan Jenis Gangguan Perkembangan Pada Anak. *Media Informatika*, Vol. 6, No. 1.
- [4] Sutojo, T., Edi Mulyanto, dan Vincent Suhartono. 2011. *Kecerdasan Buatan*. Yogyakarta : Penerbit Andik.
- [5] Aribowo, A.S dan Khomsah, S.. 2011. Sistem Pakar Dengan Beberapa *Knowledge Base* Menggunakan *Probabilitas Bayes* dan Mesin Inferensi *Forward Chaining*. *Semnas IF*. 51 – 58