

KLASIFIKASI CITRA *RONTGEN* PARU-PARU DENGAN EKSTRAKSI FITUR HISTOGRAM DAN METODE *NAIVE BAYES*

Alfiana Ramdhan, M. Arief Bustomi, Faridawati.

Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember Surabaya

Jl. Arief Rahman Hakim, Surabaya 60111

E-mail: a_bustomi@physics.its.ac.id

Abstrak— Klasifikasi citra foto *Rontgen* paru-paru dengan ekstraksi fitur histogram dan metode *Naive Bayes Classifier* telah dilakukan. Data yang digunakan dalam penelitian ini adalah berupa data citra grayscale foto *Rontgen* paru-paru berjumlah 150 data dengan rincian 90 data untuk pelatihan dan 60 data untuk pengujian. Pengolahan citra (*pre-processing image*) dilakukan untuk meningkatkan kualitas visual citra sehingga dapat mempermudah proses klasifikasi citra. Dalam proses ini, citra masukkan difilter dengan filter median untuk menghilangkan noise dalam citra. Citra yang telah difilter kemudian dilakukan pemerataan atau ekualisasi histogram. Selanjutnya, dilakukan ekstraksi nilai-nilai fitur histogram yang ada di dalam citra tersebut. Nilai fitur histogram yang digunakan dalam penelitian ini ialah rerata, standart deviasi, varians, skewness, kurtosis, beda nilai aras maksimum dan rerata, serta entropy. Pada proses pelatihan didapatkan nilai akurasi sebesar 100%. Pada proses pengujian didapatkan bahwa metode *Naive Bayes Classifier* dapat mengenali pola masukan citra dan dapat mengklasifikasikan citra. Jika dibandingkan dengan hasil penelitian sebelumnya menggunakan jaringan syaraf tiruan *backpropagation*, pada penelitian ini terjadi peningkatan akurasi yakni sebesar 73%.

Kata Kunci : Citra Foto *Rontgen*, *Naive Bayes Classifier*, Pengolahan Citra

I. PENDAHULUAN

Dewasa ini perkembangan teknologi pengolahan citra berkembang dengan pesat. Perkembangan aplikasi dalam dunia medis dan space program membuat teknik pengolahan citra digital digunakan dalam aplikasi berskala besar. Prosedur komputer yang digunakan untuk meningkatkan kontras atau kode level intensitas ke dalam warna mempermudah interpretasi X-ray (sinar X) dan citra lain yang digunakan dalam dunia industri, medis, dan ilmu biologi[5]. Dalam bidang medis, pengolahan citra dapat diaplikasikan untuk membantu dokter dalam mendiagnosa suatu penyakit.

Paru-paru merupakan salah satu organ tubuh yang sulit dideteksi dan didiagnosis oleh kebanyakan ahli radiolog. Metode bidang kesehatan yang digunakan untuk menangkap tampilan kondisi organ paru-paru salah satunya dengan menggunakan teknologi sinar-X (foto *Rontgen*). Hasil proses dari sinar-X memberikan citra yang berbeda antara paru-paru yang sehat dan yang tidak sehat, seperti paru-paru normal ataupun kanker paru-paru. Namun, pemeriksaan kanker paru-paru dari citra hasil foto *Rontgen* masih memiliki kekurangan yaitu beberapa praktisi medis seperti dokter-dokter spesialis paru-paru masih mengandalkan pengamatan visual dalam

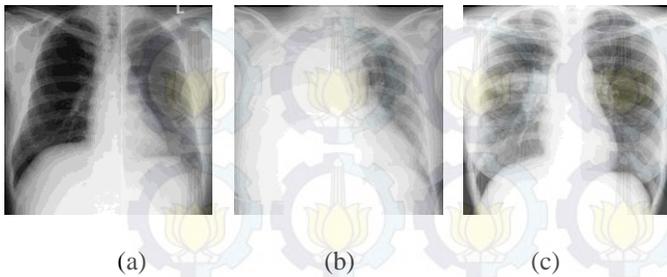
pembacaan hasil foto *Rontgen* sehingga hasilnya sangat subjektif. Oleh karena itu diperlukan perangkat lunak yang mampu mendeteksi kanker paru-paru sebagai pembanding dari kerja para praktisi medis, sehingga perangkat lunak ini dapat membantu keakuratan penentuan deteksi kanker paru-paru[2].

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu obyek. *Naive Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian [4].

Dalam penelitian ini, metode klasifikasi yang digunakan adalah *Naive Bayes Classifier*. Kelebihan *Naive Bayes Classifier* (NBC) adalah sederhana tetapi memiliki akurasi yang tinggi. Algoritma NBC yang sederhana dan kecepatannya yang tinggi dalam proses pelatihan dan klasifikasi membuat algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi[8]. Metode ini merupakan pengembangan dari metode *Bayesian Classification*. *Bayesian Classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*.

Pada penelitian ini citra paru-paru akan diklasifikasikan ke tiga kondisi, yakni kondisi normal, kanker, dan efusi yang dibedakan berdasarkan warna pada paru-paru. Paru-paru normal akan tampak hitam dan tidak terdapat adanya flek putih yang menutupi paru-paru. Paru-paru yang terkena kanker akan terlihat adanya flek-flek putih yang menutupi paru-paru. Sedangkan untuk efusi warna putih yang menutupi paru-paru lebih rapat dan dominan [6].

Pengolahan citra digital merupakan suatu kajian ilmu yang mengolah data berupa gambar dengan bantuan sistem teknologi computer, baik tentang perbaikan kualitas citra ataupun pengenalan citra, semua dapat di pelajari dalam ilmu ini. Pengolahan citra bertujuan memperbaiki kualitas citra agar mudah diinterpretasikan oleh manusia atau computer. Dalam penelitian ini, pengolahan citra yang digunakan adalah median filter dan ekualisasi histogram adaptif. Filter median sangat terkenal karena untuk jenis tertentu dari random noise, filter median memberikan kemampuan pengurangan noise pada ukuran yang sama[5]. Sedangkan Ekualisasi histogram merupakan salah satu bagian penting dari beberapa aplikasi pengolahan citra. Tujuan dari teknik ini adalah untuk menghasilkan histogram citra yang seragam. Teknik ini dapat



Gambar 1. Contoh citra digital hasil *rontgen* paru-paru: (a) normal, (b) efusi, (c) kanker

dilakukan pada keseluruhan citra atau pada beberapa bagian citra saja.

Pada dasarnya ekualisasi histogram adaptif sama dengan ekualisasi histogram biasa, hanya saja pada ekualisasi histogram adaptif citra dibagi menjadi blok-blok (*sub-image*) dengan ukuran $n \times n$. Kemudian pada setiap blok dilakukan proses ekualisasi histogram [1]. Penelitian ini dilakukan untuk mengetahui keakuratan dari sistem perangkat lunak dalam mengklasifikasikan citra hasil *Rontgen* paru-paru menggunakan metode *Naïve Bayes Classifier* dengan mengacu pada penelitian sebelumnya dengan judul “Klasifikasi Citra Paru-paru dengan Ekstraksi Fitur Histogram dan Jaringan Syaraf Tiruan *Backpropagation*”. Diharapkan penelitian ini akan menghasilkan performansi yang lebih optimal dari penelitian sebelumnya.

II. METODE

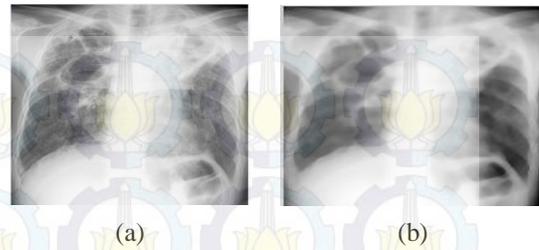
A. Obyek penelitian

Obyek yang digunakan dalam penelitian ini adalah citra *Rontgen* paru-paru berbentuk *softcopy* dari mesin X-Ray digital sebanyak 150 citra yang telah didiagnosa oleh dokter sebelumnya. Obyek tersebut dibagi untuk keperluan pelatihan perangkat lunak sebanyak 90 sampel dengan rincian 30 sampel citra X-ray paru-paru normal, 30 sampel citra X-ray paru-paru yang terkena kanker dan 30 sampel citra X-ray paru-paru yang terkena efusi dan untuk keperluan pengujian perangkat lunak sebanyak 60 sampel dengan rincian 20 sampel citra X-ray paru-paru normal, 20 sampel citra X-ray paru-paru yang terkena kanker dan 20 sampel citra X-ray paru-paru yang terkena efusi. Sedangkan dalam perancangan perangkat lunak digunakan pemrograman Matlab versi 7.10.0.499(R2010a).

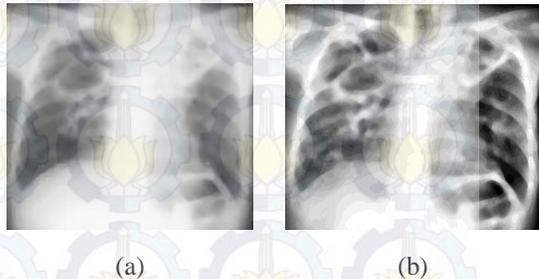
B. Pemrosesan Awal Citra

Pemrosesan awal citra bertujuan untuk meningkatkan kualitas dari citra dan menonjolkan informasi yang terkandung di dalam citra sehingga dapat mempermudah proses yang akan dilakukan selanjutnya. Pemrosesan citra awal yang dilakukan adalah pengubahan tipe citra dari tipe RGB menjadi tipe *grayscale* atau skala keabuan. Tujuan dari konversi tipe citra ini untuk mempermudah analisis citra lebih lanjut.

Kemudian dilakukan proses filterisasi citra untuk mengurangi noise pada citra dengan menggunakan median filter dan proses ekualisasi histogram adaptif untuk memperlihatkan detail pada citra. Citra hasil pemrosesan awal dapat dilihat pada gambar 2 (a)



Gambar 2. (a) Citra sebelum filtering, (b) Citra hasil filtering



Gambar 3. (a) Citra sebelum proses Ekualisasi, (b) Citra sesudah proses Ekualisasi

Pada Gambar 2 dapat dilihat bahwa citra hasil median filter masih memiliki tingkat kekontrasan yang rendah dan detail paru-paru masih kurang jelas. Oleh sebab itu, diperlukan proses perbaikan citra menggunakan ekualisasi histogram adaptif (*Adaptive Histogram Equalization*). Melalui proses tersebut dapat dilihat bahwa citra hasil ekualisasi histogram adaptif memiliki nilai keabuan yang lebih merata pada setiap tingkatan keabuan sehingga detail dari citra tiap kelompok kelas semakin terlihat.

C. Ekstraksi Fitur Citra

Setelah tahap pemrosesan awal, maka proses selanjutnya adalah tahap ekstraksi fitur. Ekstraksi fitur adalah proses untuk memunculkan ciri dan mereduksi dimensi citra dari dimensi tinggi ke dimensi yang lebih rendah. Proses ini bertujuan untuk mendapatkan informasi-informasi penting dari citra paru-paru. Citra yang telah difilter kemudian diekstrak nilai-nilai fitur histogram yang ada pada citra tersebut. Fitur histogram yang digunakan pada penelitian ini adalah nilai rata-rata (*mean*), nilai standar *deviasi*, nilai *kurtosis*, nilai *skewness* dan nilai *entropy* [5].

D. Klasifikasi Citra dengan Naive Bayes Classifier

Metode yang digunakan untuk pengklasifikasian pada penelitian ini adalah metode *Naive Bayes Classifier* (NBC). Metode ini menghitung nilai peluang terbesar. Pada proses ini dilakukan dua tahap yakni proses pelatihan dengan NBC dan proses pengujian. Pelatihan dengan *Naive Bayes Classifier* merupakan proses pelatihan yaitu menghitung *mean* dan *deviation standart* tiap fitur tiap kelas. Pengujian dengan *Naive Bayes Classifier* merupakan proses pengujian yaitu menghitung nilai peluang terbesar dengan menggunakan persamaan.2 yakni probabilitas distribusi gaussian. Pada proses ini akan diklasifikasikan citra sesuai pada kelasnya.

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut :

$$P(C | F_1, \dots, F_n) = \frac{P(C) P(F_1, \dots, F_n | C)}{P(F_1, \dots, F_n)} \dots\dots\dots(1)$$

Dimana variabel *C* mempresentasikan kelas, sementara variabel $F_1 \dots F_n$ mempresentasikan karakteristik-karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel dengan karakteristik tertentu dalam kelas *C* (*posterior*) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas *C* (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus (3) dapat pula ditulis secara sederhana sebagai berikut [4]:

$$Posterior = \frac{prior \times likelihood}{evidence} \dots\dots\dots(2)$$

Analisa proses klasifikasi dilakukan dengan menggunakan performansi sistem jaringan. Performansi sistem jaringan diuji dengan menggunakan nilai akurasi. Dimana nilai akurasi merupakan tingkat keakuratan sistem dalam mengenali data input yang diberikan sehingga menghasilkan output yang benar. Secara matematis dapat dirumuskan sebagai berikut:

$$akurasi = \frac{jumlah\ data\ yang\ benar}{jumlah\ data\ keseluruhan} \times 100\% \dots\dots\dots(3)$$

III. HASIL DAN PEMBAHASAN

a. Pelatihan Sistem Jaringan

Metode *Naive Bayes Classifier* merupakan pengklasifikasian sederhana yang menggunakan teori peluang dengan rumus turunan dari *teorema bayes*. Pada proses pengklasifikasian ini terdapat dua proses, yakni proses pelatihan citra dan proses pengujian citra.

Proses awal ini merupakan proses pelatihan dengan menghitung nilai *mean* dan *variant* tiap fitur tiap kelas. Pada tahap pelatihan ini, digunakan 90 data pelatihan yang dibagi menjadi 30 citra hasil *Rontgen* paru-paru dengan keadaan *efusi*, 30 citra hasil *Rontgen* paru-paru dengan keadaan *kanker* dan 30 citra hasil *Rontgen* paru-paru dengan keadaan *normal*. Saat proses pelatihan dijalankan maka data hasil ekstraksi fitur dapat dimasukkan secara otomatis. Berdasarkan data yang telah diinputkan diperoleh nilai *mean* dan *variant* untuk setiap kelas fitur yang selanjutnya digunakan untuk mencari peluang kelas terhadap data citra uji. Performansi dari sistem dapat dilihat dari perhitungan tingkat akurasi yang menunjukkan tingkat akurasi dari sistem dalam mengenali input yang diberikan. Dari pelatihan tersebut diperoleh nilai akurasi pelatihan seperti pada tabel 4.1 berikut.

Tabel 1. Hasil klasifikasi citra latih

Citra Latih	Terdeteksi sebagai			Tingkat Akurasi	Jumlah Citra
	Normal	Kanker	Efusi		
Normal	30	0	0	100%	30
Kanker	0	30	0	100%	30
Efusi	0	0	30	100%	30
Rata-rata prosentase				100%	90

Hasil dari proses pelatihan menunjukkan tingkat akurasi sebesar 100% pada masing-masing kelas.

b. Pengujian Sistem Jaringan

Proses ini merupakan pengujian yaitu menghitung nilai peluang dengan menggunakan rumus probabilitas distribusi gaussian, sesuai dengan persamaan berikut.

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \left[-\frac{1}{2} \left(\frac{(x_i - \mu_{ij})^2}{\sigma_{ij}^2} \right) \right] \dots\dots\dots(4)$$

Pada distribusi ini dikarakteristikan pada dua parameter input yaitu *mean* (μ) dan *varian* (σ^2) untuk tiap kelas fitur yang dicari dan peluang kelas, kemudian mencari peluang terbesar. Proses ini menggunakan 60 data citra uji. Dari pengujian tersebut, diperoleh hasil klasifikasi citra uji seperti pada tabel 2 berikut.

Tabel 2. Hasil klasifikasi citra uji

Citra Latih	Terdeteksi sebagai			Tingkat Akurasi	Jumlah Citra
	Normal	Kanker	Efusi		
Normal	16	2	2	80%	20
Kanker	11	9	0	55%	20
Efusi	0	3	17	85%	20
Rata-rata prosentase				73%	60

Asumsi yang diambil dalam pengklasifikasian citra ini adalah fitur berupa data numerik dipandang sebagai kumpulan fitur yang saling bebas (*independen*) dan proses penentuan kelas dari data uji dilakukan dengan cara menghitung besarnya probabilitas terpilihnya data uji dengan karakteristik tertentu dalam kelas tertentu (*Posterior*) dan probabilitas munculnya kelas tertentu (*Prior*) [2].

Dari tabel 2 dapat dikatakan bahwa sistem masih belum bisa mengklasifikasikan citra dengan maksimal dimana nilai akurasi pengujian sebesar 70%. Sistem lebih banyak mengenali jenis citra paru-paru efusi dan lebih sedikit mengenali jenis citra paru-paru kanker. Hal ini disebabkan karena dari 60 data citra uji sebanyak 17 data memiliki nilai probabilitas yang lebih tinggi pada kelas efusi dibandingkan pada kelas lain dan sisanya sebanyak 3 data uji masuk ke dalam kelas kanker. Begitupula pada kelas kanker, sistem lebih sedikit mengenali jenis citra ini karena dari hasil perhitungan hanya sebanyak 9 data uji yang memiliki nilai probabilitas tinggi di dalam kelas kanker. Selain itu,

banyaknya jumlah data latih juga berpengaruh pada peningkatan akurasi karena data latih merupakan data yang akan dijadikan acuan untuk membuat model klasifikasi. Oleh sebab itu, pada penelitian ini sistem masih belum bisa mengklasifikasikan citra uji dengan tingkat akurasi 100% sempurna.

Jika dibandingkan dengan penelitian yang dilakukan sebelumnya oleh Hasan Bisri pada tahun 2013, penelitian ini dilakukan dengan bahan yang sama namun dengan metode yang berbeda. Pada penelitian sebelumnya didapatkan nilai akurasi pelatihan sebesar 98,89% dan nilai akurasi pada pengujian sebesar 65%. Namun, pada penelitian ini menunjukkan nilai akurasi pada proses pelatihan dan proses pengujian mengalami peningkatan daripada penelitian sebelumnya. Bahkan pada proses pelatihan, jaringan dapat mengenali seluruh citra masukkan dengan ditandai dengan nilai akurasi pelatihan mencapai 100%. Sedangkan pada proses pengujian, sistem dapat mengenali citra dengan nilai akurasi mencapai 73% karena metode yang digunakan dapat mengestimasi nilai berdasarkan nilai probabilitas dari setiap kejadian yang muncul serta parameter parameter yang diinputkan masing masing bersipat independen atau tidak ada hubungan satu dengan yang lainnya. Hal ini membuktikan bahwa metode *Naive Bayes Classifier* dapat digunakan untuk memprediksi dan mengklasifikasikan citra paru-paru ke dalam kelas *efusi*, kanker dan normal.

IV. KESIMPULAN

Beberapa hal yang dapat disimpulkan dari penelitian yang telah dilakukan adalah metode *Naive Bayes Classifier* dapat digunakan untuk memprediksi dan mengklasifikasikan citra paru-paru ke dalam kelas *efusi*, kanker dan normal. Jika dibandingkan dengan hasil penelitian sebelumnya menggunakan jaringan syaraf tiruan backpropagation, pada penelitian ini terjadi peningkatan akurasi yakni sebesar 70%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ibu Faridawati selaku dosen pembimbing yang telah memberi bimbingan, saran serta diskusi sehingga penelitian ini dapat terselesaikan.

DAFTAR PUSTAKA

- [1] Bisri, Hasan. 2013. *Klasifikasi Citra Paru-paru dengan Ekstraksi Fitur Histogram dan Jaringan Syaraf Tiruan Backpropagation*. J. Sains dan Seni POMITS, vol.2. 2337-3520
- [2] Dewi, Nina Awaliah. Suprpto. Afif, Ahmad. 2013. *Identifikasi Jenis Sel Leukosit pada Citra Mikroskopis menggunakan Ekstraksi Fitur Geometri dan Klasifikasi Naive Bayes*. Universitas Brawijaya. Malang
- [3] Listyana, Latifah. 2013. *Implementasi Learning Vector Quantization (LVQ) untuk Klasifikasi Kanker Paru dari Citra Foto Rontgen*. Universitas Airlangga. Surabaya
- [4] Natalius, Samuel. 2010. *Metode Naive Bayes Classification dan Penggunaannya pada Klasifikasi Dokumen*. Institut Teknologi Bandung. Bandung
- [5] Prasetyo, Eko. 2011. *Pengolahan Citra Digital dan Aplikasinya menggunakan MATLAB*. Andi. Yogyakarta
- [6] Saksono, Hanung Tyas. Rizal, Achmad. Usman, Koredianto. 2010. *Pendeteksian Kanker Paru-paru dengan menggunakan Transformasi Wavelet dan Metode Linear Discriminant Analysis*. Institut Teknologi Telkom. Bandung
- [7] Seibert, J.A. 2004. *X-Ray Imaging Physics For Nuclear Medicine Technologies*. Part I : Basic Principles of X-Ray Production. *Journal of Nuclear Medicine Technology*. 32(3) : 139-147
- [8] Uluwiyah Nur, R.A dan Kusumaningsih, Ari. *Deteksi Manusia dengan menggunakan Histogram Of Oriented Gradients dan Naive Bayes Classifier*. Universitas Trunojoyo Madura. Madura