

Ciri Statistis Histogram Citra Sinar-X Untuk Deteksi Tb Paru Dengan Komputer

Ratnasari Nur Rohmah

Jurusan Teknik Elektro
Universitas Muhammadiyah Surakarta
E-mail: Ratnasari.Rohmah@ums.ac.id

Nurokhim

Badan Tenaga Atom Nasional
E-mail: nurokhim@batan.go.id

Abstract—Penelitian ini mengembangkan metode pendeteksian penyakit TB paru dengan komputer menggunakan ciri statistis histogram citra sinar-X digital. Pendeteksian meliputi pra-proses ekualisasi histogram, proses segmentasi citra dengan pengambilan area paru berdasar *ROI template*, ekstraksi ciri statistis histogram citra dengan metode *PCA (Principle Componen Analysis)*, dan klasifikasi dengan *minimum distance classifier*. Uji coba memperlihatkan metode histogram ekualisasi dan penggunaan *ROI template* yang tepat memberi kontribusi positif dalam kalkulasi ciri statistis histogram citra. Penerapan transformasi *PCA* dalam ekstraksi ciri dan dikombinasikan dengan fungsi pengklasifikasi *minimum distance classifier* memperlihatkan hasil uji pendeteksian yang cukup baik. Pada uji coba dengan data citra uji primer, pada penggunaan jarak Mahalanobis akurasi pendeteksian mencapai 95,455%, sedangkan pada penggunaan jarak Euclidean, dengan akurasi mencapai 94,318%. Meskipun demikian, hasil uji coba penerapan metoda ini pada data citra sekunder tidak memperlihatkan akurasi sebaik pada penerapan pada data citra primer. Kondisi data citra sekunder berbeda dengan citra primer menjadi penyebab perbedaan akurasi tersebut, dan menjadi peluang penelitian lanjutan untuk mengatasi perbedaan tersebut.

Kata-kata kunci: TB paru, citra sinar-X, ciri statistik

I. PENDAHULUAN

TB adalah penyakit akibat infeksi *Mycobacterium tuberculosis* yang dapat mengenai hampir semua organ tubuh dengan lokasi terbanyak di paru sebagai lokasi infeksi primer [1]. Jumlah kasus TB di Indonesia pada tahun 2011 mencapai 235 per 100.000 penduduk [2]. Sedangkan untuk wilayah Daerah Istimewa Yogyakarta, tercatat 2.419 kasus TBC, dengan 1220 merupakan kasus baru [2]. Angka ini cukup memprihatinkan, jika dilihat dari kenyataan bahwa penderita TB dengan usia produktif 25 – 64 tahun merupakan jumlah penderita terbesar dengan prosentase 71,31%.

Penegakan diagnosis TB oleh dokter radiolog salah satunya dilakukan dengan pengamatan visual pada citra sinar-X toraks untuk melihat ada atau tidaknya gambaran radiologis sugestif TB [1]. Pengamatan visual radiolog ini memiliki faktor

subyektifitas yang besar. Satu gambaran radiologis pada suatu citra sinar-X mungkin akan ditafsirkan berlainan oleh radiolog yang berbeda. Selain sifat subyektifitas, cara manual ini menjadi kendala saat rasio jumlah radiolog dengan pasien tidak mencukupi [3].

Meskipun di rumah sakit telah tersedia data citra digital, namun deteksi TB masih dilakukan secara manual berdasar pengamatan visual dokter atas citra sinar-X. Tujuan penelitian ini adalah untuk mengatasi sifat subyektifitas dan kendala tidak seimbangnya rasio radiolog dan pasien, dengan membuat sistem pendeteksian penyakit TB paru berdasar citra sinar-X dengan bantuan komputer. Pendeteksian dilakukan berdasar ciri citra sinar-X hasil komputasi komputer, sehingga diharapkan hasil pendeteksian yang lebih obyektif.

Citra sinar-X yang digunakan oleh dokter dalam penegakan diagnosis ini merupakan citra *greyscale*, sehingga ciri tekstural menjadi ciri dominan dari data citra ini [4]. Tekstur sangat penting dalam karakterisasi jaringan dan pengenalan struktur patologis [5]. Ciri tekstural citra yang didasarkan pada ciri statistis ini akan menghasilkan jumlah ciri yang tidak banyak namun merupakan ciri yang relevan dan dapat membedakan kondisi tekstural citra [6]. Hasil penelitian pendahuluan ini memperlihatkan kemungkinan digunakannya ciri statistik citra sebagai dasar pendeteksian citra.

Pendeteksian penyakit TB paru pada penelitian ini meliputi pra-proses perbaikan kualitas data citra, proses segmentasi citra dengan pengambilan area paru, ekstraksi ciri, dan klasifikasi citra atas citra TB dan citra non-TB. Ekstraksi ciri dilakukan atas ciri statistis histogram citra terukur dengan metode transformasi *PCA*. Transformasi *PCA* dilakukan selain untuk memilih ciri dengan mereduksi dimensi ciri, juga untuk menghilangkan redundansi dari ciri yang terukur [7]. Ciri terukur umumnya akan mengandung redundansi dimana ciri saling berkorelasi sehingga perlu ekstraksi ciri untuk menghilangkan korelasi ini. Ciri hasil ekstraksi merupakan ciri 'khas' yang digunakan sebagai dasar klasifikasi citra TB dan non-TB dengan pengklasifikasi *minimum distance classifier*.

II. ALAT DAN BAHAN

Data primer dalam penelitian ini adalah 138 buah foto Rontgen (citra sinar-X) toraks digital yang di ambil di bagian radiologi RSUP Sardjito, Yogyakarta. 50 data citra digunakan sebagai citra referensi dalam pengembangan sistem pendeteksian, dan 88 citra digunakan untuk citra uji coba. Pengambilan data dilakukan dengan metode sampling kebetulan. Data citra berupa citra grayscale dengan format 'bmp' dan merupakan citra berbingkai dengan ukuran citra 3520×3250 piksel untuk citra non-TB, dan ukuran citra 3520×3520 piksel, 3520×4280 piksel, dan 4280×3520 piksel, untuk citra TB. Validasi data primer ini didasarkan pada hasil diagnosa dokter yang ada dalam data tulisan yang menyertai data citra.

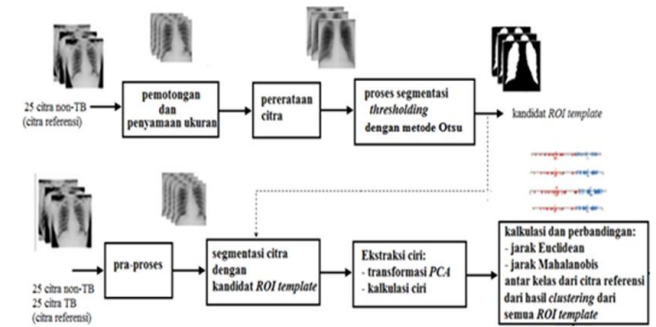
Data sekunder dalam penelitian ini adalah data yang digunakan oleh peneliti Jaeger untuk uji coba pendeteksian TB berdasar citra sinar-X. Data diperoleh atas kerja sama dengan Dr. Sameer K Antani, berupa 58 citra TB dan 80 citra normal dari Montgomery County (MC) dataset [8][9]. Ukuran data sekunder adalah: 4020 x 4892 dan 4892 x 4020. Validasi data sekunder ini didasarkan catatan yang menyertai data. Penelitian ini dilaksanakan dengan menggunakan perangkat komputer dengan prosessor Intel(R) Celeron(R) CPU 877 @1,40 GHz, dan memori (RAM) 4 GB.

III. METODA PENELITIAN

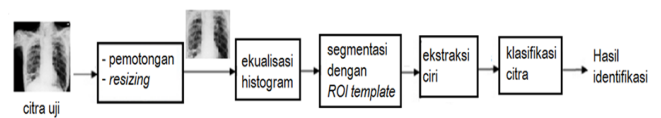
Langkah-langkah pengembangan metode pendeteksian dalam penelitian ini dibagi dalam dua tahap, tahap pra-deteksi dan tahap pendeteksian. Tahapan pra-deteksi, seperti terlihat pada Gambar 1, meliputi pembuatan *ROI template* dan penemuan ciri 'khas' untuk citra referensi TB dan citra referensi non-TB. Pembuatan *ROI template* dilakukan dengan proses *thresholding* atas rerata 25 citra referensi non-TB. Rerata citra referensi merupakan rerata seluruh citra referensi non-TB, setelah penyamaan ukuran dengan penskalaan citra dan *up-sampling* dengan interpolasi bikubik. Penggunaan citra rerata ini dilakukan untuk mengatasi ketidak-seragaman dalam ukuran dan posisi thoraks dalam citra sinar-X. Obyek yang akan dipisahkan dari latar belakang merupakan obyek tunggal yaitu area paru, sehingga teknik *thresholding* yang digunakan adalah *bi-level thresholding* [10].

Penemuan ciri 'khas' dengan transformasi *PCA* dilakukan atas lima ciri statistis histogram citra: rerata, std (standar deviation), skewness, kurtosis, dan entropi. Kalkulasi ciri dilakukan pada 25 citra referensi TB dan 25 citra referensi non-TB yang telah mengalami peningkatan kualitas dengan ekualisasi histogram, dan segmentasi citra dengan pengambilan area citra sesuai *ROI template*. Peningkatan kualitas citra dengan ekualisasi histogram digunakan dalam

penelitian ini dilakukan untuk mengatasi ketidakseragaman dalam kualitas kecerahan data citra.



Gambar 1. Langkah-langkah pada tahapan pra-deteksi yang menghasilkan *ROI template*, konstanta transformasi *PCA*, dan parameter kelas TB dan kelas normal



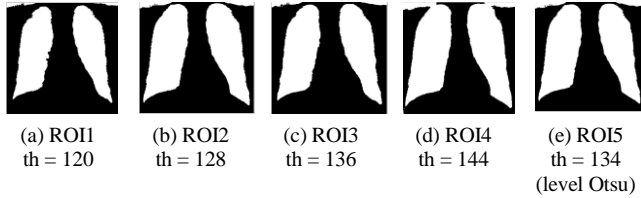
Gambar 2. Tahapan proses deteksi satu citra uji

Tahap pendeteksian citra merupakan tahap pendeteksian citra uji dengan parameter citra referensi. Gambar 2 memperlihatkan langkah-langkah dalam tahapan deteksi yang meliputi: pemotongan dan penyamaan ukuran citra uji, segmentasi citra dengan *ROI template*, kalkulasi ciri dan ekstraksi ciri dengan *PCA*, dan klasifikasi citra. Pemotongan citra dilakukan untuk menghilangkan bingkai citra dan penyamaan ukuran dilakukan sesuai dengan ukuran *ROI template* yang akan digunakan. Transformasi *PCA* pada citra uji dilakukan dengan menggunakan konstanta-konstanta transformasi yang diperoleh pada tahapan pra-deteksi. Klasifikasi citra uji sebagai citra TB atau citra non-TB dengan metode *minimum distance classifier* dengan dua tipe jarak *Euclidean* dan *Mahalanobis*. Konstanta fungsi pengklasifikasi ini merupakan parameter kelas citra referensi yang diperoleh pada tahapan pra-deteksi.

IV. HASIL DAN PEMBAHASAN

A. Pembuatan *ROI template*

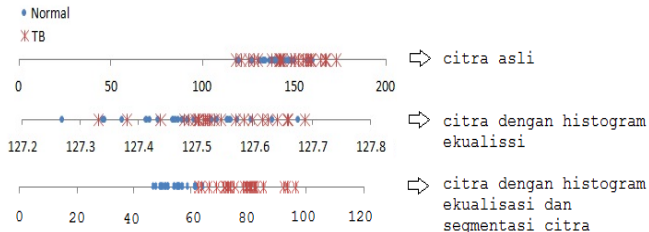
Pada penelitian ini dibuat lima *ROI templates* seperti terlihat pada Gambar 3. Penelitian ini menggunakan empat nilai ambang di sekitar nilai tengah jangkauan level intensitas, yaitu: $th = 120$, $th = 128$, $th = 136$, and $th = 144$. Selain keempat *template* tersebut, proses *thresholding* dilakukan juga dengan menggunakan level *threshold* Otsu dari citra rerata yang digunakan dalam eksperimen. Level *threshold* Otsu yang diperoleh dari rerata 25 citra normal referensi, $th = 134$, mempunyai nilai yang hanya sedikit saja berbeda dengan level ambang perkiraan pada ROI3.



Gambar 3. Lima *ROI template* hasil *thresholding* pada rerata citra referensi kelas normal dengan berbagai nilai ambang

B. Pengaruh ekualisasi histogram dan segmentasi citra pada pengelompokan kelas citra referensi berdasar ciri statistis.

Penelitian ini mengkalkulasi ciri statistis histogram citra orde pertama. Lima ciri statistis yang digunakan pada penelitian ini adalah: rerata, deviasi standar (*std*), skewness, kurtosis, dan entropy. Eksperimen yang dilakukan pada penelitian ini memperlihatkan pengaruh ekualisasi histogram dan segmentasi citra dengan *ROI template* memperlihatkan pengaruh positif pada proses ekstraksi ciri citra.

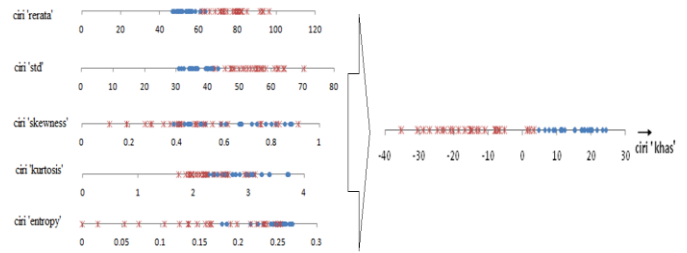


Gambar 4. Pengelompokan 50 citra referensi berdasar ciri statistis 'rerata' yang dikalkulasi dari: (a) citra asli; (b) citra setelah histogram ekualisasi; dan (c) citra setelah histogram ekualisasi dan segmentasi citra.

Gambar 4 memperlihatkan pengelompokan citra TB dan non-TB dari 50 citra referensi yang didasarkan pada ciri statistis histogram 'rerata'. Terlihat pengaruh proses segmentasi yang sangat besar pada pengelompokan citra berdasar citra asli. Hasil ini memperlihatkan pentingnya ketepatan dalam pengambilan area paru dari seluruh frame citra. Sedangkan pengaruh peningkatan kualitas citra dengan ekualisasi histogram tidak memperlihatkan pengaruh yang signifikan, hal ini dikarenakan kualitas data citra primer yang cukup seragam.

C. Ekstraksi ciri dengan Principle Component Analysis

Penelitian ini menggunakan ciri tekstural untuk digunakan sebagai dasar dalam mengklasifikasikan citra sebagai TB atau non-TB. Kalkulasi ciri tekstural dilakukan dengan mengukur lima ciri statistis histogram citra pada *ROI* citra. Ekstraksi ciri selanjutnya dilakukan dengan transformasi *PCA* untuk mendapatkan satu 'ciri khas' dari kelima ciri statistis yang terukur. Gambar 5 memperlihatkan pengelompokan citra berdasar 'ciri khas' hasil ekstraksi ciri dengan *PCA*, yang lebih terpisah, dimana tidak ada tumpang-tindih citra citra referensi antara citra kelas TB dan non-TB.



Gambar 5. Pengelompokan kelas citra TB dan non-TB berdasar masing-masing ciri terukur (kiri) dan pengelompokan kelas citra berdasar 'ciri khas' hasil ekstraksi ciri dengan *PCA* (kanan).

D. Deteksi citra uji

Uji coba metode yang dikembangkan pada penelitian ini dilakukan dengan mendeteksi baik data citra primer maupun sekunder. Data primer yang digunakan dalam pengujian adalah 88 data citra primer yang terdiri dari 41 citra TB dan 47 citra non-TB. Sedangkan data sekunder yang digunakan dalam citra uji adalah 33 citra sekunder TB dan 55 citra sekunder non-TB. Keseluruhan data citra uji disertai dengan catatan hasil diagnosa dokter yang digunakan sebagai standar kebenaran pada perhitungan akurasi hasil deteksi. Uji coba dilakukan dengan berbagai *ROI template* dan berbagai ukuran citra.

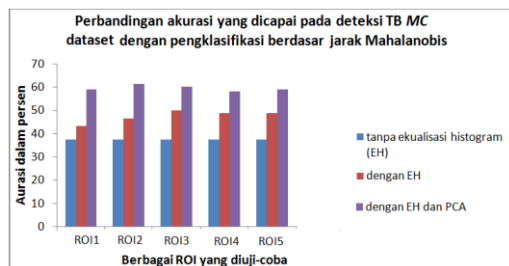
Uji coba pengaruh penggunaan berbagai ukuran *template* ternyata memperlihatkan bahwa akurasi sama sekali tidak dipengaruhi oleh penskalaan citra dengan interpolasi bikubik. Sedangkan pengaruh bentuk *template* mempengaruhi akurasi pendeteksian citra uji primer. Tabel 1 memperlihatkan hasil uji coba memperlihatkan *ROI template* yang dibuat dengan proses *thresholding* dengan ambang $th = 128, 134, \text{ dan } 136$, memperlihatkan hasil akurasi terbaik pada pendeteksian citra dengan *minimum Mahalanobis distance classifier*. Sedangkan pada klasifikasi yang didasarkan pada *minimum Euclidean distance classifier*, akurasi hampir sama pada semua *template*, kecuali pada *template* dengan $th = 144$, yang memperlihatkan akurasi lebih rendah. Uji coba memperlihatkan penggunaan jarak Mahalanobis memberikan hasil yang lebih baik, dengan akurasi tertinggi 95,455%, dibandingkan penggunaan jarak Euclidean, dengan akurasi tertinggi 94,318%.

Tabel 1 Hasil uji coba deteksi citra uji (data primer) berdasar satu ciri hasil transformasi *PCA* dari berbagai penggunaan *ROI template*

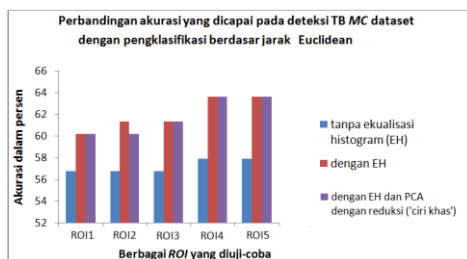
ROI template	Accuracy (%)		Sensitivity (%)		Specitivity (%)	
	Euclidean	Mahalanobis	Euclidean	Mahalanobis	Euclidean	Mahalanobis
ROI1	94,318	94,318	90,244	100,000	97,872	89,362
ROI2	94,318	95,455	90,244	100,000	97,872	91,489
ROI3	94,318	95,455	92,683	97,561	95,745	93,617
ROI4	93,182	93,182	92,683	97,561	93,617	89,362
ROI5	94,318	95,455	92,683	97,561	95,745	93,617

Metode yang dikembangkan pada penelitian ini juga diuji coba pada data sekunder *MC Dataset*. Pengujian dilakukan dengan melakukan prosedur yang sama, baik pada tahap pra-deteksi maupun tahap deteksi pada data sekunder *MC Dataset*. Pengujian akurasi uji coba pada data citra sekunder dengan

penggunaan jarak Euclidean memberikan hasil akurasi 63,64%. Sedangkan akurasi yang dicapai dengan penggunaan jarak Mahalanobis sebesar 60,2%. Jika dilihat dari perbandingan pendeteksian citra atas citra TB dan non-TB, akurasi yang didapatkan pada citra sekunder ini jauh di bawah akurasi yang didapatkan pada citra primer. Perbedaan hasil ini bisa disebabkan oleh kondisi data yang berbeda. Perbedaan histogram citra pada data primer dan sekunder, seperti pada Gambar 8, memperlihatkan data sekunder sudah mengalami pra-proses sebelum diterima peneliti. Salah satu proses yang terlihat dari pengamatan adalah adanya proses penghalusan (*smoothing*) yang tidak dilakukan pada data primer.

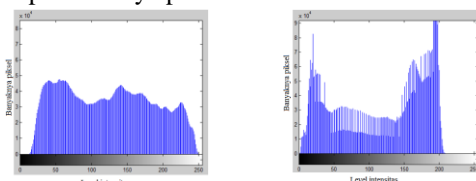


Gambar 6. Grafik yang memperlihatkan perbandingan dari akurasi deteksi citra sekunder berdasar lima ciri statistis dengan *minimum Mahalanobis distance classifier*



Gambar 7. Grafik yang memperlihatkan perbandingan dari akurasi deteksi citra sekunder berdasar lima ciri statistis dengan *minimum Euclidean distance classifier*

Perbedaan lain adalah data primer citra TB yang digunakan hampir semua merupakan citra TB lanjut, sedangkan pada data sekunder, sebagian data citra TB yang ada adalah citra TB dini. Hasil ini menunjukkan bahwa metode yang diusulkan oleh peneliti dalam penelitian ini memberikan hasil baik dengan keterbatasan: TB yang dideteksi adalah TB paru tahap lanjut, dan harus dari sumber data yang sama. Keterbatasan ini memberikan peluang untuk penelitian bagi pengembangan sistem pendeteksian TB tahap dini dengan metode yang tidak terpengaruh pada adanya perbedaan kualitas data citra.



Gambar 8. Perbedaan yang terlihat pada histogram citra-citra sekunder (kiri) dan citra-citra primer (kanan)

V. KESIMPULAN

Hasil penelitian ini memperlihatkan bahwa ciri statistis histogram citra sinar-X dapat digunakan sebagai dasar pendeteksian penyakit TB dengan 15 komputer. Metode pendeteksian yang melibatkan pra-proses ekualisasi histogram, segmentasi citra dengan *ROI template*, ekstraksi ciri statistis dengan transformasi *PCA*, dan penggunaan *minimum distance classifier* menghasilkan akurasi tertinggi 95,455%. Meskipun hasil uji coba pada data sekunder tidak memperlihatkan akurasi setinggi pada uji coba data primer, namun penggunaan ekualisasi histogram dan *ROI template* memperlihatkan kontribusi positif dalam pemisahan kelas citra TB dan non-TB, baik pada uji coba data primer maupun data sekunder. Perbedaan akurasi pada data primer dan sekunder terkait dengan perbedaan kondisi data primer dan sekunder. Keterbatasan ini memberikan peluang untuk penelitian bagi pengembangan sistem pendeteksian TB yang bisa mengatasi adanya perbedaan kondisi data citra.

REFERENCES

- [1] Puspongoro, Hardiono D., "Standar Pelayanan Medis Kesehatan Anak", first ed., Badan Penerbit IDAI, Indonesia, (2005).
- [2] Depkes, 2011. "Strategi Nasional Pengendalian TB di Indonesia 2010 – 2014".
- [3] Research and Development Team Health Professional Education Quality (HPEQ) Project. "Potret Ketersediaan Dan Kebutuhan Tenaga Dokter", Ditjen Dikti Kemendikbud, 2010.
- [4] Kachanubal, T., Udomhunsakul, S. 2008, "Rock Textures Classification Based on Textural and Spectral Features. World Academy of Science", *Engineering and Technology* 39, (2008) 110 – 116.
- [5] Mitrea, D., S.Nedevschi, M.Lupsor, R.Badea, "Exploring the Textural Parameters obtained from Ultrasound Images for Modeling the Liver Pathological Stages in the Evolution towards Hepatocellular Carcinoma", In *IEEE Conference of Automation, Quality and Testing, Robotics*, vol. 3, no., pp.128-133, 22-25 May. 2008
- [6] Aggarwal, N.; R. K. Agrawal, "First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images," *Journal of Signal and Information Processing*, 2012, 3, 146-153.
- [7] Partridge, M., & Calvo, R. A. 1998. "Fast dimensionality reduction and simple PCA", *Intelligent Data Analysis*, 2(1-4), 1998, 203-214.
- [8] Candemir S, Jaeger S, Musco J, Xue Z, Karagyris A, Antani SK, Thoma GR, Palaniappan K., "Lung Segmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration". In *IEEE Transactions on Medical Imaging*, Vol. 33, No.2, pp. 577 – 590, February, 2014..
- [9] Jaeger S, Karagyris A, Candemir s, Folio L, Siegelman J, Callaghan FM, Xue Z, Palaniappan K, Singh RK Antani SK. "Automated tuberculosis screening using chest radiograph". In *IEEE Transactions on Medical Imaging*, Vol. 33, No.2, pp. 233 – 245, February, 2014..
- [10] Gonzalez, R. C. and R. E. Woods, "Digital Image Processing", third ed., Pearson Education, New Jersey, 2008.