

## **Content Based Image Retrieval Using Euclidean Distance to Detect Pneumonia in X-Ray Thorax Image**

Content based image retrieval menggunakan Euclidean Distance Untuk mendeteksi Penyakit Pneumonia Pada Citra X- Ray Thorax

**Bambang Yuwono<sup>1\*</sup>, Wilis Kaswidjanti<sup>2</sup>, Nisa'ul Azizah<sup>3</sup>, Nur Heri Cahyana<sup>4</sup>**

<sup>1,2,3,4</sup> Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

<sup>1\*</sup> bambangy@upnyk.ac.id, <sup>2</sup> wilisk@upnyk.ac.id, <sup>3</sup> nisa.azizah@gmail.com,

<sup>4</sup> dsnurheri@gmail.com

\*: *Penulis korespondensi (corresponding author)*

**Keywords:** *Pneumonia, Content Based Image Retrieval Euclidean Distance, x-ray thorax*

**Kata kunci:** *Pneumonia, Content Based Image Retrieval Euclidean Distance, x-ray thorax*

### **Abstrak**

**Tujuan:** Penelitian ini bertujuan menerapkan *Content Based Image Retrieval* untuk membuat aplikasi pendeteksi Pneumonia atau normal pada citra x-ray thorax menggunakan metode *euclidean distance* serta mengetahui tingkat akurasi dari aplikasi dalam menentukan apakah pneumonia atau normal.

**Perancangan/metode/pendekatan:** *content-based image retrieval* (CBIR) adalah metode pencarian gambar yang menganalisis fitur seperti warna, bentuk, tekstur, atau informasi lain yang dapat diekstraksi dari suatu gambar untuk menemukan gambar dalam database gambar yang besar. CBIR dibagi menjadi tiga tahap: ekstraksi fitur dari gambar, penyimpanan fitur, dan pencarian gambar. Masukkan query citra x-ray ke dalam aplikasi, dan citra query x-ray ini akan melalui proses ekstraksi fitur warna dan ekstraksi fitur tekstur untuk menemukan kemiripan dengan citra di database. Proses penemuan citra membandingkan jarak *Euclidean* citra query dengan citra dalam database.

**Hasil:** Hasil pengujian dengan pengukuran kemiripan *Euclidean Distance* diperoleh akurasi sebesar 96% pada data latih 100 dan data uji 22, dengan jumlah citra yang ditampilkan 4. Sedangkan bila pengujian menggunakan data yang sudah dilatihkan menghasilkan akurasi 100%.

### **State of the art:**

Dengan terdapatnya aplikasi pendeteksi pneumonia maka bisa memudahkan membaca hasil gambar x- ray thorax.

## 1. Pendahuluan

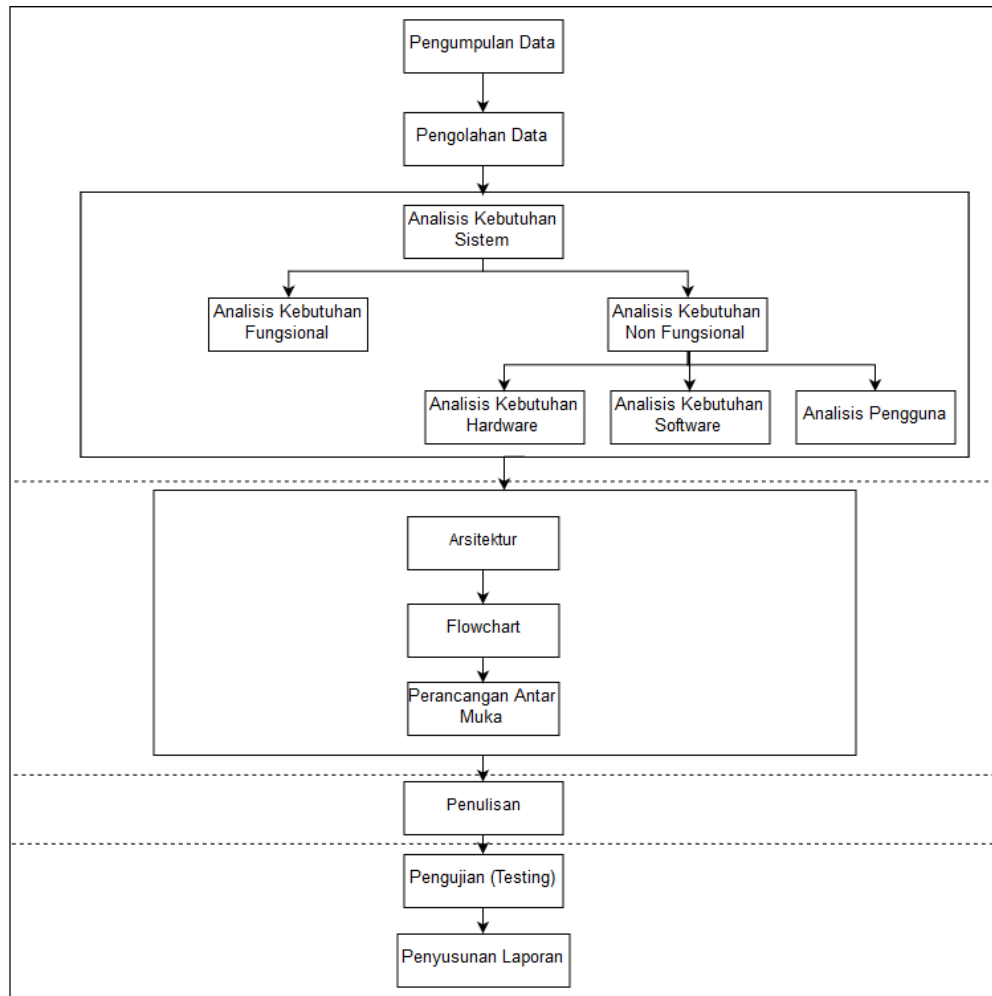
Penggunaan *x-ray thorax* menjadi salah satu pemeriksaan radiologis yang paling umum dalam mendiagnosa penyakit jantung dan paru-paru. Pembacaan *x-ray thorax* mengandalkan pengetahuan dari profesional dan proses observasi manual yang sangat cermat, dikarenakan patologi yang kompleks dan perubahan tekstur halus dari lesi paru yang berbeda pada gambar, ahli radiologi dapat membuat kesalahan bahkan ketika mereka telah menjalani pelatihan klinis jangka panjang dan bimbingan profesional [1]

Seringkali membutuhkan waktu lama bagi tenaga medis untuk membaca foto rontgen dada karena setiap area tidak tergambar dengan jelas. Area yang tidak terlihat jelas karena film yang terlalu terang tampak lebih gelap dari biasanya. Dapat dipastikan bahwa film tersebut kurang terang ketika tidak dapat terlihat tulang belakang di dada. Hasil rontgen dada kurang bagus, disarankan untuk mengambil foto dada lagi, sehingga membutuhkan biaya mahal serta waktu lama. Berlandaskan perihal tersebut diperlukan sesuatu aplikasi untuk mempermudah petugas kedokteran membaca hasil x- ray yang kurang jelas.

Pencarian citra berdasarkan isi citra disebut dengan *Content Based Image Retrieval (CBIR)*. CBIR merupakan salah satu metode pencarian gambar dari database gambar yang besar dengan menganalisa fitur-fitur seperti warna, bentuk, tekstur, atau informasi lainnya yang bisa di ekstrak dari gambar [2]. Terdapat 3 tahap fundamental pada CBIR yaitu tahap ekstraksi fitur dari gambar, penyimpanan fitur dan pencarian gambar [3]. Metode CBIR dengan metode *euclidean distance* dapat digunakan untuk mendeteksi apakah x-ray thorax tersebut normal atau pneumonia

Pada Penelitian [4] ini menggunakan metode analisis berbasis atlas, yang bertujuan untuk mengidentifikasi populasi dari data hasil untuk menentukan diagnosis selanjutnya. Penelitian oleh [5] berjudul Klasifikasi Penyakit Paru-Paru Berbasis Radiografi Menggunakan Metode Segmentasi Sobel. Penelitian ini dapat mengklasifikasikan 6 (enam) jenis penyakit paru dengan membandingkan piksel antara hasil deteksi tepi terhadap region paru. Penelitian yang dilakukan oleh [6] Penelitian ini menggunakan Learning vector quantization dengan akurasi data training terbaik sebesar 89,714% dan akurasi data test yang terbaik sebesar 74,00% yang didapat pada pengujian dengan learning rate=0,4. Penelitian yang dilakukan oleh [7] Penelitian ini menggunakan GLCM dan Neural Network untuk klasifikasi kualitas kematangan wortel berdasarkan tekstru kulit menghasilkan akurasi 72,50%. Penelitian yang dilakukan oleh [8] Penelitian ini menggunakan GLCM serta KNN untuk identifikasi Jenis Anggrek menghasilkan akurasi rata-rata 77%. Penelitian yang dilakukan oleh [9] Penelitian ini memanfaatkan GLCM citra buah jeruk keprok untuk klasifikasi mutu menghasilkan akurasi yang terbaik 82,5% dari jumlah data latih hingga 20, dengan jarak distance=2 pada sudut GLCM 45°.

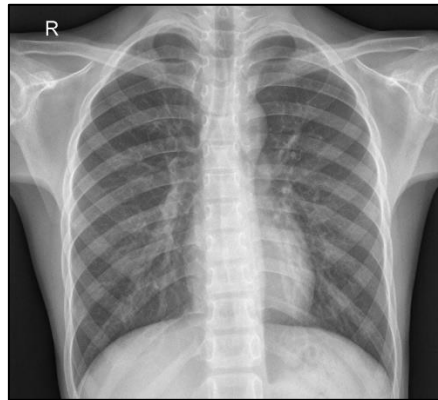
## 2. Metode/Perancangan



Gambar 1. Metodologi / Tahapan Penelitian

### 2.1. Pengumpulan Data

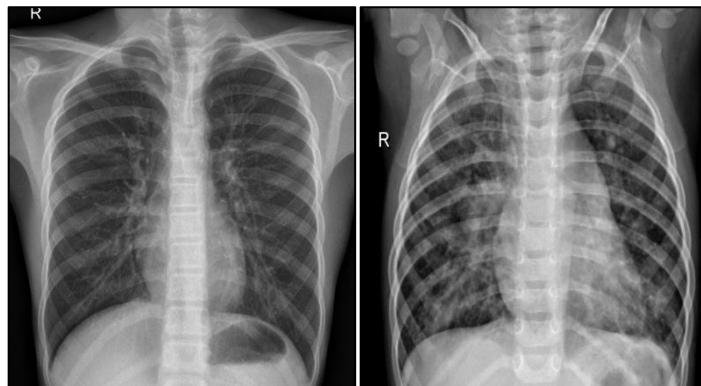
Data Primer penelitiannya berupa foto *x-ray thorax* yang dikonversi dalam file bertipe *jpeg*. Data primer berasal dari Rumah Sakit Khusus Paru (RSKP) Respira Bantul. Adapun contoh gambar *x-ray thorax* terlihat pada gambar 2.



Gambar 2. Contoh data *x-ray thorax*

## 2.2. Pengolahan Data

Selain menentukan derajat keabuan pada citra thorax, penelitian ini menggunakan tekstur sebagai parameter untuk menentukan penyakit.



Gambar 3. (Kiri) Citra *x-ray thorax* Normal (Kanan) Pneumonia

Adanya infiltrat di paru-paru pada foto rontgen dada membedakan antara paru normal dan pneumonia. Infiltrat adalah bintik atau bintik di paru-paru yang dapat dilihat pada foto rontgen dada. Adanya infiltrat menunjukkan adanya dahak (lendir) di paru-paru yang menyebabkan peradangan. Peradangan tersebut memberikan ciri-ciri eksternal seperti batuk dan demam tinggi, namun untuk memastikan lebih lanjut diperlukan citra *x-ray thorax*.

Infiltrat sendiri dalam citra *x-ray thorax* berwarna lebih terang daripada paru-paru normal. Pada paru-paru normal bagian yang dekat dengan tulang rusuk bagian tengah mempunyai bentuk yang tajam dan tidak banyak bagian putih. Sedangkan paru-paru yang terdeteksi pneumonia mempunyai bintik putih yang cukup banyak dan bentuknya cukup lebar sehingga terlihat lebih besar.

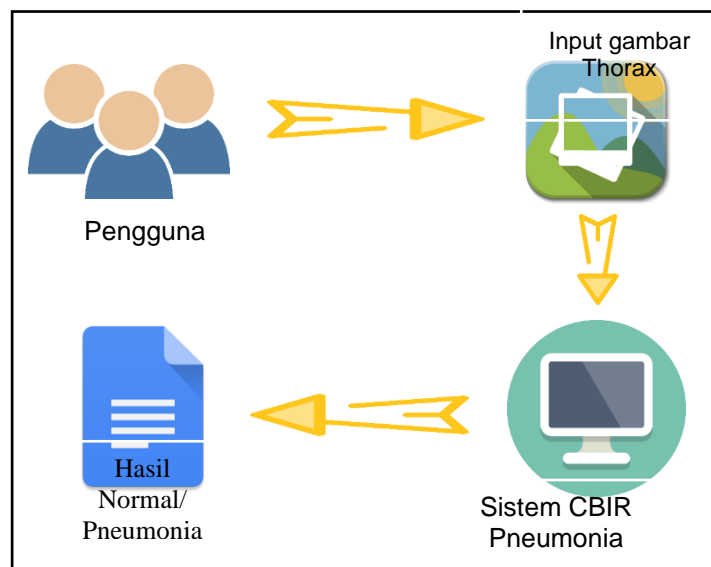
Untuk pengolahan data citra *x-ray thorax* dilakukan bersama dengan pakar radiolog dan dokter spesialis penyakit paru-paru RSKP Respira Bantul. Pakar radiolog dan dokter bergantian memberi sedikit pengetahuan kepada penulis untuk membedakan citra thorax yang normal dan *pneumonia*. Data citra yang didapat dari rumah sakit respirasi ada sejumlah 120 citra thorax yang kemudian dibagi menjadi beberapa bagian untuk keperluan penelitian.

Tabel 1. Foto x-ray yang digunakan

No	Nama Data	Jenis Data	Jumlah Data
1	foto x-Ray thorax Normal	Database	50
2	foto x-Ray thorax Pneumonia	Database	50
3	foto x-Ray thorax Normal	Data Uji	11
4	foto xRay thorax Pneumonia	Data Uji	11

### 2.3. Arsitektur Sistem

Arsitektur sistem menggambarkan proses yang sedang digunakan. Arsitektur sistem dari *content based image retrieval* ini dapat ditunjukkan pada gambar 4.



Gambar 4. Arsitektur sistem

Pengguna akan memasukkan citra thorax untuk diolah ke dalam arsitektur sistem. Sistem kemudian akan mengolah citra tersebut dengan menggunakan prosedur-prosedur yang terdapat di dalamnya, seperti pembacaan database, ekstraksi fitur, serta perhitungan persamaan fitur, sehingga dihasilkan keluaran yang menunjukkan apakah gambar itu terdeteksi pneumonia atau tidak.

### 2.4. Euclidean Distance

Euclidean distance merupakan algoritma untuk menghitung jarak antara dua objek, yaitu dengan cara menghitung selisih jarak antara dua buah vektor yang akan dibandingkan seperti pada rumus (1) yang akan digunakan untuk pengenalan objek [10] [11]

Menghitung akar kuadrat selisih dua buah vektor, sehingga semakin kecil jaraknya maka vektor masukan dan vektor referensi akan semakin mirip dengan hasil yang diperoleh.

$$d = \sqrt{\sum_{k=1}^m (f_{d_{i,k}} - k_j)^2} \quad (1)$$

Keterangan :

$d$  = jarak euclidean

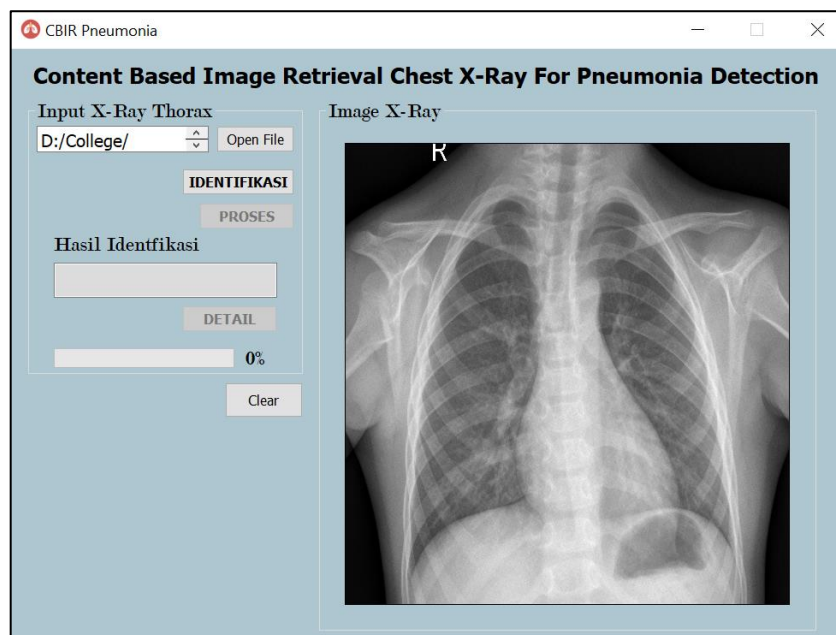
$fd_{i,k}$  = data training

$k_j$  = data uji

$M$  = jumlah data

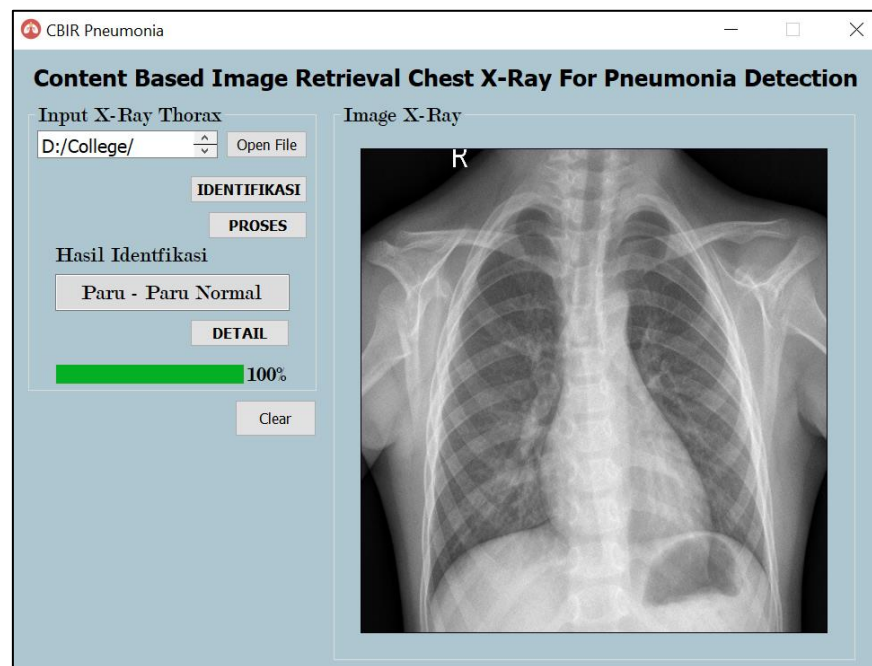
### 3. Hasil dan Pembahasan

Gambar 5 merupakan tampilan halaman *Form* Utama setelah menjalankan proses pada tombol *Open File*. Proses pertama yang dilakukan adalah membaca folder untuk meng-*input* citra *x-ray thorax*. Lalu membaca citra yang telah dipilih dan ditampilkan pada *Display Image*. Setelah membaca Citra *x-ray thorax* yang di *open* akan muncul ke dalam *GroupBox Image X-Ray*.



**Gambar 5.** Tampilan *Form* setelah melakukan proses *Open File*

Tombol identifikasi digunakan untuk membaca *database* yang merupakan kumpulan dari citra training yang nantinya akan dibandingkan dengan citra *x-ray thorax input* dengan metode perbandingan. Setelah membaca gambar pada *database* selanjutnya nilai vektor masing-masing citra yang ada di *database* dihitung. Setelah itu dihitung nilai terdekat dengan citra *input* menggunakan *Euclidean Distances* dibandingkan dengan citra training. Setelah menemukan jarak terdekat dengan *euclidean distances* maka akan ditentukan apakah citra input merupakan paru-paru normal atau *Pneumonia*. Gambar 6 merupakan tampilan *Form* hasil Identifikasi berdasarkan *Input X-ray Thorax* yang di Proses. Pada gambar tersebut terlihat hasil identifikasi menghasilkan paru-paru normal.



**Gambar 6.** Form Identifikasi

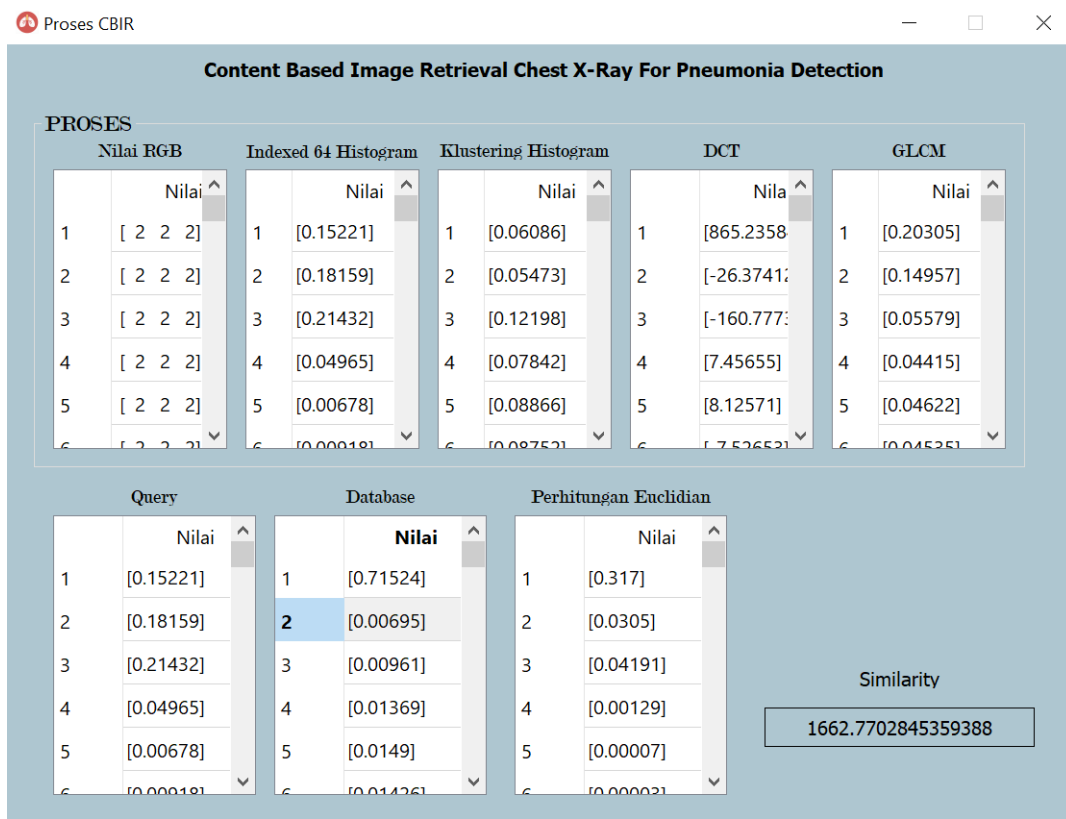
Tombol Proses digunakan untuk menampilkan jalannya proses yang ada pada aplikasi seperti yang terlihat pada gambar 7. Pada tombol proses akan dimunculkan nilai RGB citra input yang kemudian diproses pada *color histogram*, *discrete cosine transform* dan *gray level cooccurrence matrix* yang dimunculkan merupakan hasil angka dari perhitungan tiap metode dari citra yang telah diinputkan.

Histogram warna digunakan untuk mengekstrak warna dengan menghitung nilai distribusi warna pada citra uji dan citra latih dengan menghitung jumlah piksel dari setiap bagian rentang warna. *Discrete cosine transform* (DCT) merupakan teknik konversi sinyal yang memecah beberapa sinyal menjadi komponen frekuensi dasar. *Discrete cosine transform* (DCT) dapat memampatkan gambar semaksimal mungkin. Menggunakan teknik kompresi *lossy* pada DCT, kompresi gambar yang dihasilkan sedikit mengurangi warna (piksel) tetapi tampak identik dengan gambar asli sebelum kompresi.

Sehingga setelah citra uji melalui proses *color histogram* untuk mendapatkan hasil distribusi warna sesuai pixelnya masing-masing, maka selanjutnya akan masuk ke dalam proses kompresi gambar dengan metode *discrete cosine transform* (DCT). Pada metode ini kompresi citra digunakan agar warna (pixel) citra uji dapat disesuaikan dengan citra training yang sebelumnya sudah di kompresi.

GLCM adalah matriks skala keabuan yang mewakili hubungan antara satu tingkat keabuan dengan tingkat keabuan lainnya. Metode ini sering diterapkan untuk analisis suatu piksel pada gambar dan mengetahui derajat keabuan yang terjadi. Setelah citra uji melalui proses *color histogram* dan DCT maka selanjutnya adalah masuk pada penilai kesamaan tingkat keabuan pada nilai pixel dengan menggunakan metode GLCM. Dimana apabila hasil akhir GLCM sudah muncul akan dilakukan perhitungan dengan metode *euclidean distances* dimana apabila hasil semakin sedikit maka citra uji semakin mirip dengan citra training.

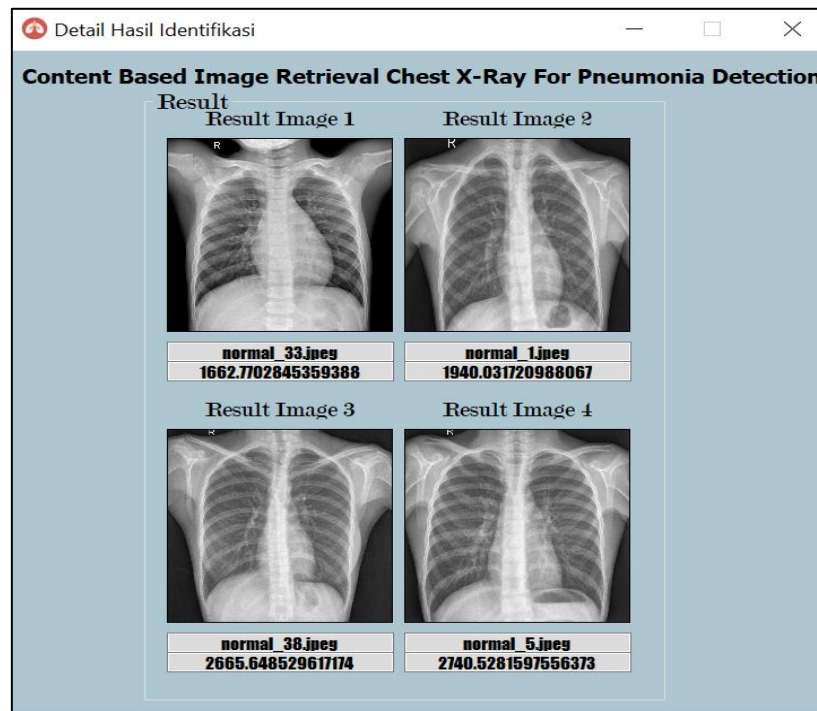
Sehingga tidak ada batasan terhadap nilai apakah citra uji pneumonia atau tidak. Karena pengambil keputusan untuk penentuan apakah citra uji pneumonia atau tidak adalah kesamaan antara nilai pixel dari citra uji dan citra training.



Gambar 7. Tampilan setelah Tombol Proses

Gambar 8 merupakan tampilan pada halaman detail dengan memunculkan 4 gambar citra x-ray thorax yang mempunyai perhitungan nilai *euclidean distances* yang paling mirip dengan citra *input*. Paling mirip disini ditunjukkan dengan perhitungan nilai *euclidean* dimana semakin kecil nilai *euclidean* maka semakin mirip citra. Metode ini digunakan dengan perhitungan jarak dari *image retrieval* untuk membantu proses klasifikasi pada citra training dan citra uji. Nilai perhitungan *euclidean distances* muncul pada *textbox* dibawah citra *x-ray thorax* yang mempunyai nilai vektor paling mirip dengan citra *input*.





Gambar 7. Tampilan Tombol Detail

Aplikasi Deteksi pneumonia dengan CBIR ini diuji menggunakan dua data uji yaitu citra uji yang ada di database dan citra uji di luar database. Aplikasi ini menggunakan algoritma *euclidean distance* untuk menghitung nilai kemiripannya. Nilai pada gambar database yang mendekati nol maka adalah gambar yang paling mirip dengan gambar query sehingga nilai yang paling tidak mendekati angka nol merupakan gambar yang paling tidak mirip.

Pengujian pertama menggunakan data uji yang diambil dari sebagian data latih (data training). Data uji yang digunakan 20 data uji dari keseluruhan data training sebanyak 100 data. Hasil pengujian yang pertama menunjukkan bahwa untuk seluruh data uji, seluruhnya dapat dikenali dengan benar, atau tingkat akurasi pengenalannya sebesar 100%. Hal ini membuktikan bahwa aplikasi dapat mengenali citra x-ray thorax yang sudah dimasukkan ke dalam data latih sebelumnya dengan sangat baik.

Pengujian kedua menggunakan 22 citra x-ray thorax dengan perincian 11 data normal dan 11 data pneumonia yang belum ada dalam latih. Tabel 3 menunjukkan tingkat akurasinya sebesar 96%, dimana dari 22 data uji ada satu data yang salah dalam mendeteksi.

Tabel 3. Pengujian Validasi Data

No	Nama Citra Uji	Hasil Pakar	Citra Kenal pada Sistem	Euclidean	Keterangan
1	Normal_1.jpeg	Normal	Normal_39	0,1285	Benar
2	Normal_2.jpeg	Normal	Normal_33	0,1663	Benar
3	Normal_3.jpeg	Normal	Normal_33	0,1663	Benar
4	Normal_4.jpeg	Normal	Normal_9	0,3554	Benar
5	Normal_5.jpeg	Normal	Normal_16	0,1554	Benar
6	Normal_6.jpeg	Normal	Normal_28	0,1215	Benar
7	Normal_7.jpeg	Normal	Normal_6	0,2098	Benar
8	Normal_8.jpeg	Normal	Normal_33	0,1773	Benar
9	Normal_9.jpeg	Normal	Normal_46	0,1599	Benar
10	Normal_10.jpeg	Normal	Normal_29	0,893	Benar
11	Pneumonia_11.jpeg	Pneumonia	Pneumonia_47	0,793	Benar
12	Pneumonia_12.jpeg	Pneumonia	Pneumonia_10	0,2686	Benar
13	Pneumonia_13.jpeg	Pneumonia	Pneumonia_10	0,2686	Benar
14	Pneumonia_14.jpeg	Pneumonia	Pneumonia_9	0,920	Benar
15	Pneumonia_15.jpeg	Pneumonia	Pneumonia_9	0,841	Benar
16	Pneumonia_16.jpeg	Pneumonia	Pneumonia_28	0,1385	Benar
17	<b>Pneumonia_17.jpeg</b>	<b>Pneumonia</b>	<b>Normal_49</b>	<b>0,1804</b>	<b>Salah</b>
18	Pneumonia_18.jpeg	Pneumonia	Pneumonia_18	0,1503	Benar
19	Pneumonia_19.jpeg	Pneumonia	Pneumonia_28	0,1758	Benar
20	Pneumonia_20.jpeg	Pneumonia	Pneumonia_36	0,2436	Benar
22	Pneumonia_21.jpeg	Pneumonia	Pneumonia_40	0,2346	Benar
21	Normal_11.jpeg	Normal	Normal_40	0,1882	Benar

#### 4. Kesimpulan

Berdasarkan hasil dan pembahasan dari penelitian ini, maka dapat disimpulkan sebagai berikut:

1. Penggunaan *Content Based Image Retrieval* dapat menentukan gambar rontgen dada normal atau ada pneumonia
2. Metode pencarian *Euclidean Distance* dapat digunakan untuk mengukur jarak terdekat antara nilai vektor gambar uji dengan gambar pada database.
3. Hasil pengujian dengan pengukuran kemiripan *Euclidean Distance* diperoleh akurasi sebesar 96% pada data latih 100 dan data uji 22, dengan jumlah citra yang ditampilkan 4. Sedangkan bila pengujian menggunakan data yang sudah dilatihkan menghasilkan akurasi 100%.

#### Daftar Pustaka

- [1] Guan, Q. *et al*, "Thorax disease classification with attention guided convolutional neural network", *Pattern Recognition Letters*, 131, pp. 38–45. doi: 10.1016/j.patrec.2019.11.040
- [2] R.S. Torres dan A.X Falcao, *Content-Based Image Retrieval; Theory dan Application*, Institute of Computing, State University of Campinas, Campinas, SP, Brazil, 2006
- [3] B. Singh dan W. Ahmad, *Content-Based Image Retrieval: A Review Paper*, *Internasional Journal of Computer Science and Mobile Computing*, volume 3 issue 5, 2014

- 
- [4] A. V. Faria, K. Oishi, S. Yoshida, A. Hillis, M. I. Miller and S. Mori, "Content-Based Image Retrieval fro Brain MRI: An Image-Searching Engine and Population-Based Analysis to Utilize past Clinical Data for Future Diagnosis," *NeuroImage. Clinical*, vol. 7, p. 367–376, 2015.
- [5] R. Rahmadewi and R. Kurnia, "Klasifikasi Penyakit Paru Berdasarkan Citra Rontgen dengan Metoda Segmentasi Sobel," *Jurnal Nasional Teknik Elektro*, vol. 5, no. 1, pp. 7-12,
- [6] Jimmi Lin, Hafiz Irsyad, "Klasifikasi Pneumonia Pada Citra X-Rays Paru-paru Menggunakan GLCM dan LVQ. Jurnal Algoritme, vol.1, No.2 , pp. 184-194, 2021
- [7] M rinandar Tasya, Bambang Soedjiono, Emha Taufik Luthfi, "Klasifikasi Kualitas Kematangan Worter Menggunakan Metode GLCM dan Neural Network, Junal Fateksa, vol 5 No. 2 , pp 1-10, 2020
- [8] Dinar Putra Pamungkas, "Ekstraksi Citra Menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek, jurnal INNOVATICS, vol.1 No. 2, pp 51-56, 2019
- [9] Restu Widodo, Agus Wahyu Widodo, Arry Supriyanto, "Pemanfaatan Ciri Gray Level Co-Occurance Matrix (GLCM) Citra Buah Jeruk Keprok (*Citrus Reticulata Blanco*) untuk Klasifikasi Mutu, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol.2 No. 11, pp 5769-5776, 2018
- [10] Tessa S Warongan, Sherwin R.U.A Sompie, Agustinus Jacobus, "Penerapan Metode Content-Based Image Retrieval untuk Pengenalan Jenis Bunga", *Jurnal Teknik Informatika*, vol 13, no.3, pp 2301-8364, 2018
- [11] R. Wulanningrum and A. Rachmad, "Pengenalan Rumput Laut Menggunakan Euclidean Distance Berbasis Ekstraksi Fitur," in *Seminar Nasional Aplikasi Teknologi Informasi 2012 (SNATI 2012)*, Yogyakarta, 2012.