

**KLASIFIKASI ANOMALI PADA GAMBAR RONTGEN DADA DENGAN
METODE *MACHINE LEARNING* MENGGUNAKAN *HISTOGRAM OF
ORIENTED GRADIENT* DAN *RANDOM FOREST***

SKRIPSI

EKA WULANDARI

171402084



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA**

MEDAN

2024

KLASIFIKASI ANOMALI PADA GAMBAR *RONTGEN* DADA DENGAN
METODE *MACHINE LEARNING* MENGGUNAKAN *HISTOGRAM OF ORIENTED*
GRADIENT DAN *RANDOM FOREST*

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Komputer Teknologi Informasi

EKA WULANDARI

171402084



PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024

PERSETUJUAN

Judul : KLASIFIKASI ANOMALI PADA GAMBAR
RONTGEN DADA DENGAN METODE MACHINE
LEARNING MENGGUNAKAN HISTOGRAM OF
ORIENTED GRADIENT DAN RANDOM FOREST

Kategori : Skripsi

Nama Mahasiswa : Eka Wulandari

Nomor Induk Mahasiswa : 171402084

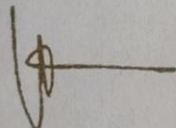
Program Studi : Sarjana (S-1) Teknologi Informasi

Fakultas : Ilmu Komputer dan Teknologi Informasi
Universitas Sumatera Utara

Medan, 21 Mei 2024

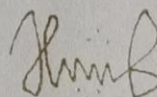
Komisi Pembimbing:

Pembimbing 2



Ulfi Andayani S.Kom., M.Kom.
NIP. 198604192015042004

Pembimbing 1



Niskarto Zendrato S.Kom., M.Kom
NIP. 198909192018051001

Diketahui/disetujui oleh
Program Studi S-1 Teknologi Informasi
Ketua,

Dedy Arisandi, ST, M.Kom
NIP. 197908312009121002



PERNYATAAN

KLASIFIKASI ANOMALI PADA GAMBAR *RONTGEN* DADA DENGAN
METODE *MACHINE LEARNING* MENGGUNAKAN *HISTOGRAM OF ORIENTED*
GRADIENT DAN *RANDOM FOREST*

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 20 Mei 2024



EKA WULANDARI

171402084

UCAPAN TERIMA KASIH

Alhamdulillah, rasa syukur saya ucapkan kepada Allah yang Maha Pengasih lagi Penyayang, yang sudah menciptakan semua makhluk yang ada dilangit dan dibumi baik itu yang terjangkau oleh mata manusia ataupun tidak terjangkau, yang sudah memberikan begitu banyak nikmat yang besar kepada semua makhluk dan sudah sepantasnya sebagai hamba yang lemah memohon petunjuk dan jalan terbaik kepada-Nya atas izin Allah lah penulis dapat menuntaskan Tugas Akhir ini. Disini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dan begitu berperan dalam peroses penyelesaian skripsi ini.

Dengan kerendahan hati, penulis meminta maaf apabila pembaca menemukan kekurangan didalam penelitian yang sudah saya susun ini, dan dengan hati terbuka juga saya menerima atas masukan yang membangun dari para pembaca.

Dalam penyelesaian skripsi ini, penulis ucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan dan doa kepada penulis. Oleh karena itu, penulis mengucapkan terima kasih kepada sebagian pihak, dan beberapa diantaranya yang sangat berperan penting yaitu:

1. Bapak Prof. Dr. Muryanto Aminn, S.Sos., M.Si selaku Rektor Universitas Sumatera Utara.
2. Ibu Dr.Maya Silvi Lydia, B.Sc., M.Sc selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi.
3. Bapak Dedy Arisandi, M.Kom., sebagai Ketua Prodi Departemen Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi dan Pak Ivan Jaya., M.Kom selaku Sekretaris Prodi Departemen Teknologi Informasi Universitas Sumatera Utara Periode 2021-2026.
4. Kedua orang tua penulis yang bernama Bapak Susanto dan Ibu Susanti dan ketiga saudara kandung penulis yang bernama Yolanda, Melita dan Ivan Karunia, serta seluruh keluarga besar saya.
5. Bapak Dr. Mohammad Andri Budiman selaku Dosen Pembina Organisasi UKMI Al-Khuwarizmi Fasilkom-Ti Universitas Sumatera Utara.

6. Bapak Niskarto Zendrato S.Kom., M.Kom. yang merupakan Dosen mata kuliah dan Pembimbing 1 dan Ibu Ulfi Andayani S.Kom., M.Kom. yang merupakan Dosen Pembimbing 2 dan juga dosen mata kuliah.
7. Bapak dan Ibu Dosen Penguji, atas masukan dan juga saran yang diberikan kepada penulis dalam menyusun Skripsi ini.
8. Seluruh teman-teman Organisasi UKMI Al-Khuwarizmi Fasilkom-Ti, dan seluruh UKMI yang berada di Universitas Sumatera Utara.
9. Teman dalam suka dan duka penulis yang telah memberikan semangat dan *effort* yang luar biasa dalam menyelesaikan skripsi penulis yaitu Aflah Mutsanni Pulungan S.Kom, Bagus Syahputra Tambunan S.Kom dan Ayu Laila Hutagalung S.Kom.
10. Teman yang sudah banyak membantu penulis dalam berjuang untuk meraih gelar S.Kom di Fasilkom-TI Universitas Sumatera Utara, yaitu Adelia Salmah Siregar S.Kom, Majidah Atmayana Purba, Aida Suryana Ritonga, Melati Yulvira Salsabilah S.Kom, Nuraini, Sintha Sintya Rani S. Kom, Eka Khairani S.Kom, Ayu Laila S.Kom, Lisa Ayuning Tias, Nia Geminda S.T, Bella Savira, Tria, Dita Febiyan Pratiwi S.Pd, Ahmad Adil S.Kom dan Sindy Ariani S.Kom.
11. Teman-teman kom C, serta seluruh teman angkatan 2017 yang merupakan teman berjuang dalam masa perkuliahan di Teknologi Informasi.

Muda-mudahan Allah berikan rezeki yang berlimpah serta pahala –Nya. Dan semoga skripsi ini dapat berguna dan bermanfaat untuk diri saya, untuk keluarga, untuk lingkungan saya, untuk seluruh masyarakat, untuk mahasiswa yang membutuhkan dan untuk negara.

Medan, 20 Mei 2024



Eka Wulandari

ABSTRAK

Sebelum mengonfirmasi adanya kondisi yang tidak normal, dokter menggunakan pemeriksaan radiologi sebagai langkah awal dalam mendiagnosis penyakit paru-paru. Untuk mengklasifikasi kondisi yang tidak biasa terjadi dilakukan pemeriksaan melalui *x-ray thorax*. Oleh karena itu, penelitian ini bertujuan untuk membangun suatu model dalam mengklasifikasi objek abnormal menggunakan algoritma *Histogram of Oriented Gradient* (HOG) dan *Random Forest* (RF) pada citra *x-ray thorax*, dimana akan terdapat 12 objek abnormal yang akan diklasifikasi. Diantaranya yaitu: *Atelectasis*, *Cardiomegaly*, *Concolidation*, *Infiltration*, *Nodule*, *Mass*, *Emphysema*, *Fibrosis*, *Pleural Effusion*, *Pneumothorax*, *Pneumonia* dan *No_Finding*. Ada beberapa teknik yang digunakan untuk meningkatkan akurasi model saat membangun model *training* yaitu *sqrt* dan *log2*. Hasil akurasi terbaik yang telah didapatkan dari model *Random Forest* dengan melakukan *training* menggunakan *n-estimator* sebanyak 100 dan *max features sqrt* adalah 92%, yang mana dengan hasil tersebut dapat disimpulkan bahwa metode *Histogram of Oriented Gradient* dan *Random forest* yang dipakai dalam penelitian ini dapat mengklasifikasi hasil *rontgen* dengan baik. Selain membangun model, penulis juga mengembangkan sebuah aplikasi berbasis dekstop yang bertujuan untuk membantu dokter umum dalam mempermudah proses diagnosis penyakit dengan menganalisis hasil paru. Aplikasi ini menggunakan teknologi *image processing* untuk mengklasifikasi tanda-tanda penyakit yang terlihat pada gambar *rontgen* paru-paru. Aplikasi berbasis dekstop ini diharapkan dapat membantu dokter atau petugas kesehatan terkait untuk mempermudah dalam proses diagnosis penyakit dan dapat dimanfaatkan sebagai alat pembelajaran yang baik dan berguna.

Kata kunci : Klasifikasi, *Histogram of Oriented Gradiend*, *Random Forest*, *Machine learning*, *N-estimator*, *Max Features*.

**ANOMALY CLASSIFICATION IN CHEST X-RAY IMAGES USING MACHINE
LEARNING METHOD WITH HISTOGRAM OF ORIENTED
GRADIENT AND RANDOM FOREST**

ABSTRACT

Before confirming that there is an abnormal condition, doctors use radiological examinations as the first step in diagnosing lung disease. To classify unusual conditions, an examination is carried out via a chest x-ray. Therefore, this study aims to build a model for classifying abnormal objects using the Histogram of Oriented Gradient (HOG) and Random Forest (RF) algorithms on chest x-ray images, where there will be 12 abnormal objects to be classified. Among them are: Atelectasis, Cardiomegaly, Concolidation, Infiltration, Nodule, Mass, Emphysema, Fibrosis, Pleural Effusion, Pneumothorax, Pneumonia and No_Finding. There are several techniques used to improve model accuracy when building a training model, namely sqrt and log2. The best accuracy results have been obtained from the Random Forest model by training using an n-estimator of 100 and the max features sqrt is 92%, with these results it can be concluded that the Histogram of Oriented Gradient and Random forest methods used in this study can classify X-ray results properly. In addition to building the model, the authors also developed a desktop-based application that aims to assist general practitioners in facilitating the process of diagnosing disease by analyzing lung results. This application uses image processing technology to classify signs of disease seen on X-ray images of the lungs. This desktop-based application is expected to help doctors or related health workers to simplify the process of disease diagnosis and can be utilized as a good and useful learning tool.

Keywords: *Classification, Histogram of Oriented Gradient, Random Forest, Machine learning, N-estimator, Max Features.*

DAFTAR ISI

PERSETUJUAN	ii
PERNYATAAN	iii
UCAPAN TERIMA KASIH	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
1.6. Metodologi Penelitian	5
1.7. Sistematika Penulisan	6
BAB 2 LANDASAN TEORI	7
2.1. Klasifikasi	7
2.2. Anomali	7
2.3. Akurasi	7
2.4. Gambar <i>Rontgen</i> Dada	8
BAB 3 ANALISIS DAN PERANCANGAN	33
3.1. Dataset yang digunakan	33

3.2.	Arsitektur Umum	38
3.3.	Proses Training Random forest	43
3.4.	Metode Evaluasi	44
3.5.	Perancangan Sistem	46
BAB 4 IMPLEMENTASI DAN PENGUJIAN		50
4.1.	Implementasi Sistem	50
4.2.	Implementasi Model	65
4.3.	Evaluasi	66
BAB 5 KESIMPULAN DAN SARAN		69
5.1.	Kesimpulan	69
5.2.	Saran	69
DAFTAR PUSTAKA		70

DAFTAR GAMBAR

Gambar2.1 Foto Rontgen Dada PA, AP, Lateral	8
Gambar2.2 Foto pembagian lobus pada paru-paru	9
Gambar2.3 Klasifikasi pejalan kaki	13
Gambar2.4 rumus operator sobel atau <i>centered derivatice</i>	14
Gambar2.5 Perhitungan turunan terpusat	14
Gambar2.6 Rumus Teorema Pitagoras	15
Gambar2.7 Perhitungan Nilai Magnitude	15
Gambar2.8 Proses Polarity change	16
Gambar2.9 Perhitungan cell HOG	17
Gambar2.10 Metode Pembuatan Histogram 1 (<i>Sumber: www.sefidian.com</i>)	17
Gambar2.11 Metode Pembuatan Histogram 2	18
Gambar2.12 Metode Pembuatan Histogram 3	19
Gambar2.13 Metode Pembuatan Histogram 4	19
Gambar2. 14 Nilai <i>Magnitude</i>	20
Gambar2.15 Perhitungan HOG 8 x 8 cell	21
Gambar2. 16 Normalisasi Gradien	22
Gambar2. 17 Menggabungkan 4 <i>cell</i>	22
Gambar2. 18 Hasil Fitur HOG	23
Gambar2. 19 Proses Deskriptor HOG	24
Gambar2. 20 <i>Workflow</i> dari Algoritma <i>Random forest</i>	25

DAFTAR TABEL

Tabel 2. 1 Peneliti Terdahulu	30
Tabel 3. 1 Jumlah Data	33
Tabel 3. 2 Contoh Jenis hasil Radiologi Paru	34
Tabel 3. 3 Confusion Matrix	45
Tabel 4. 1 Spesifikasi kebutuhan Hardware	50
Tabel 4. 2 Spesifikasi kebutuhan Software	50
Tabel 4. 3 Pembagian data training berdasarkan label	51
Tabel 4. 4 Pembagian data testing berdasarkan label	52
Tabel 4. 5 Implementasi Model dengan Data Testing	65
Tabel 4. 6 Nilai TP, TN, FP dan FN pada Data Testing	67

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Paru-paru adalah organ yang sangat penting bagi tubuh manusia yang berfungsi sebagai sistem pernafasan, yaitu tempat pertukaran oksigen dan karbondioksida didalam darah. Masalah yang sering terjadi pada saat bernafas adalah banyaknya udara yang dihirup dengan kualitas yang telah tercemar dan mengandung banyak bibit-bibit penyakit yang akan menyerang paru.

Penyakit paru merupakan penyakit yang berdampak sangat serius terhadap sistem pernafasan pada manusia yang dapat berakibat fatal apabila tidak segera ditangani dengan serius. Ada beberapa penyakit paru-paru yang secara umum sering dijumpai yaitu seperti, *tuberkulosis*, *bronkitis*, *pneumonia*, *kanker paru*, *emfisema* dan *pleuriti* (Rahmadewi et al., 2013). Citra *Rontgen* dada digunakan oleh dokter spesialis gangguan pernafasan untuk membantu pasien dalam menilai apakah seseorang sedang mengidap permasalahan pernafasaan atau tidak (Bambang Pulu Hartato, 2021).

Pemeriksaan penyakit paru-paru menggunakan citra hasil foto rontgen asli masih memiliki kelemahan. Dokter spesialis paru sering mengandalkan pengamatan visual untuk membaca hasil foto rontgen, yang membuat hasilnya sangat subjektif. Penelitian sebelumnya menunjukkan bahwa ahli radiologi memiliki tingkat kegagalan sebesar 30% dalam mendiagnosa bintil kecil pada paru-paru (Fontana, 1986). Dokter spesialis paru-paru harus melakukan pengamatan foto rontgen secara teliti dan mendiagnosis dengan akurat untuk mengklasifikasi kanker paru-paru pada pasien. Oleh karena itu, diperlukan perangkat lunak yang dapat mengklasifikasikan kanker paru-paru sebagai pembanding untuk kinerja praktisi medis, sehingga perangkat lunak tersebut dapat membantu meningkatkan akurasi dalam penentuan klasifikasi kanker paru-paru (Faridawati, 2015).

Menurut Departemen Kesehatan Republik Indonesia, penyakit paru-paru merupakan salah satu penyakit yang masih menjadi masalah serius hingga saat ini. Pada tahun 2009, jumlah penderita penyakit paru-paru mencapai 1,7 juta orang, sedangkan pada tahun 2010 meningkat menjadi 2,3 juta orang, dan pada tahun 2011

mencapai 4,7 juta orang. Salah satu faktor yang menyebabkan peningkatan jumlah penderita penyakit paru adalah tingginya jumlah perokok, yang mencapai 65% dari total penduduk Indonesia pada tahun 2011. Sementara itu, jumlah dokter spesialis penyakit paru-paru hanya sekitar 600-700 orang di Indonesia. Data ini menunjukkan bahwa jumlah penderita penyakit paru-paru tidak sebanding dengan jumlah dokter yang ada. Oleh karena itu, diperlukan teknologi yang dapat membantu penanganan cepat bagi penderita penyakit paru-paru. Salah satu solusinya adalah penggunaan teknologi informasi berbasis komputer (Utomo et al., 2016). Kemajuan teknologi komputer saat ini telah mendorong penelitian dan implementasi teknik pengolahan citra. Pengolahan citra memiliki peran penting dalam berbagai bidang kehidupan, termasuk penginderaan jarak jauh, militer, *Optical Character Recognition* (OCR), dan juga dalam bidang kesehatan. Contoh penggunaan pengolahan citra dalam bidang kesehatan adalah kemampuannya dalam mengklasifikasikan penyakit. Dalam konteks ini, penelitian akan membahas klasifikasi radiologi pada gambar Rontgen paru dengan tujuan membaca gambar rontgen secara akurat, yang biasanya dilakukan oleh ahli yang terampil dalam interpretasi rontgen (Rahmadewi, 2016).

Rahmat Hidayat *et al* (2021) dengan judul Penerapan Metode Pembelajaran Menggunakan Ekstraksi Fitur dan Algoritma Klasifikasi untuk Identifikasi Pengenalan Iris melakukan penelitian sebanyak 21 kali percobaan dengan membandingkan tiga algoritma ekstraksi fitur yaitu *Hu-Moments*, *Haralick* dan HOG yang dikombinasikan dengan beberapa algoritma klasifikasi seperti *Logistic Regression* (LR), *K-Nearest Neighbor* (KNN), *Linear Discriminant Analysis* (LDA), *Decision Tree Classifier* (CART), *Random forest* (RF), *Gaussian* (NB) dan *Support Vector Machine* (SVM). Berdasarkan penelitian tersebut HOG yang dikombinasikan dengan algoritma klasifikasi *Random forest* memberikan hasil evaluasi terbaik berdasarkan nilai akurasi yaitu masing-masing sebesar 80,27%. Sedangkan untuk penerapan ekstraksi fitur lainnya, penerapan pada fitur *Hu- moments* terhadap algoritma *random forest* menghasilkan nilai akurasi terbaik yaitu sebesar 43.61% dibandingkan dengan algoritma klasifikasi lainnya. Kemudian fitur *Haralick* mendapatkan nilai akurasi terbaik jika dikombinasikan dengan algoritma klasifikasi *linear discriminant analysis* dengan jumlah nilai akurasi sebesar 60.55%.

Pada penelitian yang dilakukan oleh (Mustopa *et al.*, 2022) tentang mengklasifikasi rontgen toraks dengan menggunakan *feature extraction Hu-*

Moments, *Harralic* dan Histogram untuk mencari akurasi terbaik dari hasil penklasifikasian COVID-19 dengan algoritma klasifikasi. Dalam penelitian ini juga memakai teknologi *machine learning* selama satu dekade karena telah berkembang pesat dan telah terintegrasi kedalam sistem CAD untuk memberikan akurasi yang akurat. Penelitian dilakukan dengan menguji dataset yang didapatkan dari repositori kaggle.com yang memiliki citra Sinar-X COVID-19, Sinar-X Normal, Sinar-X virus *pneumonia* dan Sinar-X bakteri-*Pneumonia*. Penelitian ini termasuk kedalam kategori baik karena menghasilkan akurasi tertinggi yaitu oleh algoritma klasifikasi *Random forest* yang mana pada hasil akurasi ini sebesar 84% dan Standar Deviasi 0.015847.

Penulis mengajukan penelitian dengan algoritma *Random forest* menggunakan HOG untuk mengekstraksi fitur objek gambar pada *rontgen* dada, karena di Indonesia penyakit paru-paru menjadi penyebab kematian tertinggi (Hreelaita Dharma Shanti, 2021). Jadi dengan dibuatnya 2 metode ini harapannya dapat membantu mempercepat pengklasifikasian data rontgen dada yang diteliti oleh para ahli.

Dengan demikian, penelitian ini akan memaparkan bagaimana performa algoritma *Random forest* yang dikombinasikan dengan *HOG* dalam pengklasifikasian penyakit paru-paru. Pengklasifikasian ini bertujuan untuk menampilkan informasi penting dari *Rontgen* dada sehingga menghasilkan akurasi terbaik dari sekumpulan data citra yang dijadikan sebagai referensi bagi penelitian selanjutnya.

Berdasarkan pada latar belakang diatas, penulis melakukan penelitian dengan judul **“Klasifikasi Anomali Pada Gambar Rontgen Dada Dengan Metode Machine learning Menggunakan Histogram of Oriented Gradient dan Random forest ”** karena masalah awal penelitian ini adalah salah satunya tentang kecepatan dan ketepatan dalam membaca *Rontgen* sehingga dibuat judul tersebut.

1.2. Rumusan Masalah

Objek abnormal yang terlihat samar menjadi masalah yang harus dihadapi oleh para dokter dalam pemeriksaan *x-ray thorax* yang menyebabkan para dokter membutuhkan ketelitian, sehingga terkendala oleh waktu. Oleh sebab itu, dibutuhkan sistem yang mampu menghasilkan akurasi tercepat untuk setiap hasil radiologi paru sehingga dapat mempercepat waktu dokter untuk menentukan jenis objek abnormal pada paru-paru.

1.3. Batasan Masalah

Batasan masalah pada penelitian ini terdapat beberapa tahapan yaitu:

1. Objek yang akan diklasifikasi adalah berjumlah 12 tipe anomali *thorax* (dada).
2. Dataset diperoleh melalui *platform* Kaggle.com yang disediakan oleh dua rumah sakit di Vietnam: Rumah Sakit 108 dan Rumah Sakit Universitas Kedokteran Hanoi.
3. Jumlah data *training* 1.920 gambar dan data *testing* 480 gambar.

1.4. Tujuan Penelitian

Tujuan penelitian ini untuk mengklasifikasikan anomali yang terjadi pada paru-paru melalui citra *Rontgen thorax* dan mempercepat hasil *Rontgen* dengan mengkombinasikan algoritma HOG (*Histogram of Oriented Gradients*) dan RF (*Random forest*).

1.5. Manfaat Penelitian

Manfaat pada penelitian ini yaitu:

1. Menambah referensi bagi pengembangan dibidang *image processing*.
2. Mengetahui kecepatan akurasi menggunakan metode *Histogram of Oriented Gradient*.
3. Diharapkan penelitian ini dapat menilai *abnormalitas X-Ray thoraks* dengan cepat berupa akurasi, sehingga mempermudah dokter umum dalam mendiagnosis penyakit paru melalui foto *X- Ray*.

1.6. Metodologi Penelitian

Berikut ini merupakan beberapa tahap yang dikerjakan penulis dalam perancangan penelitian yaitu:

1. Studi Literatur

Ditahap ini mempelajari dan mengumpulkan informasi yang dibutuhkan terkait penelitian yang nantinya akan dilakukan. Referensi ini beberapa skripsi, makalah, jurnal, buku internet, dan sumber lain yang mendukung penelitian ini dan juga berhubungan dengan algoritma *Histogram of Oriented Gradient*. .

2. Analisis Permasalahan

Permasalahan yang telah ditentukan untuk diteliti kemudian dianalisis guna memperoleh pengetahuan terkait metode yang akan diterapkan dalam melakukan prediksi yaitu menggunakan metode *random forest*.

3. Perancangan Sistem

Setelah menganalisis penulis merancang sistem dan melakukan *training* data yang nantinya berguna untuk mendapatkan bobot nilai dari data yang diteliti.

4. Implementasi Sistem

Setelah itu penulis melakukan *testing* terhadap *Random forest* dengan model yang nantinya akan dibangun, dengan menggunakan *testing data* untuk mendapatkan *output* didalam penelitian ini.

5. Pengujian Sistem

Langkah selanjutnya adalah mengklasifikasikan gambar *Rontgen* dada dengan *machine learning*, *Histogram of Oriented Gradient*, serta menguji system yang dibangun untuk mengetahui bahwa hasil dari penelitian ini sesuai dengan target.

6. Penyusunan Laporan

Dalam tahap ini merupakan tahapan terakhir, penulis akan membuat dokumentasi dan melaporkan setiap kegiatan yang dilakukan berupa laporan penelitian.

1.7. Sistematika Penulisan

Berikut sistematika penulisan yang akan menjabarkan lima bab dengan penjelasan tersendiri, dari setiap bab akan dijelaskan secara singkat sebagai berikut:

Bab 1: Pendahuluan

Bab ini memaparkan latar belakang, lalu batasan masalah, tujuan penelitian, manfaat penelitian, kemudian yang terakhir adalah metodologi penelitian dan juga sistematika penulisan.

Bab 2: Landasan Teori

Dalam bab ini berisikan tentang landasan teori untuk bisa memahami setiap permasalahan yang dibahas dan berkaitan tentang penelitian ini, seperti *Machine learning*, *HOG* dan algoritma *Random forest*.

Bab 3: Analisis dan Perancangan Sistem

Dalam bab ini menjelaskan tentang arsitektur umum dan analisis pada tahapan tentang *pre-processing*, proses *training* dan *testing* serta sampai penelitian ini mendapatkan hasil klasifikasi dengan kecepatan akurasi dari *Rontgen* dada dengan metode *Histogram of Oriented Gradient* supaya dapat digunakan untuk menguji sistem.

Bab 4: Implementasi dan Pengujian Sistem

Dalam bab 4 ini menjelaskan peng-aplikasian suatu program terhadap sistem dari suatu hasil menganalisis serta suatu perancangan yang sebelumnya telah dijelaskan pada Bab terdahulu, serta pembahasan hasil yang sudah diperoleh dari hasil pengujian sistem.

Bab 5: Kesimpulan dan Saran

Dalam bab ini berisikan tentang kesimpulan penelitian yang telah dibahas pada bab sebelumnya dan saran dari penelitian ini untuk perbaikan serta pengembangan pada penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1. Klasifikasi

Klasifikasi adalah metode untuk mengelompokkan sejumlah besar data yang diberikan berdasarkan atribut target atau keluaran yang dituju, di mana seluruh set data dapat memenuhi syarat sebagai anggota kelas. Teknik ini berguna dalam mengidentifikasi pola perilaku data. Secara singkat, klasifikasi adalah mekanisme untuk membedakan atau mengelompokkan data.

Klasifikasi *Random Forest* melibatkan pembentukan pohon dengan melatihnya menggunakan sampel pada data yang tersedia. Dalam proses klasifikasi dengan *Random Forest*, metode *voting* digunakan untuk mengambil keputusan mayoritas berdasarkan pohon-pohon yang telah terbentuk. Sebelum melakukan klasifikasi, dataset dibagi menjadi dua kategori, yaitu data untuk pelatihan dan data untuk pengujian. Tujuan dari pembagian data adalah untuk mengevaluasi kinerja sistem (Saparudin *et al.*, 2021).

2.2. Anomali

Anomali merujuk pada suatu keadaan yang tidak biasa atau berbeda dari keadaan normal yang umumnya terjadi di suatu area atau tempat. Anomali mempunyai 2 dimensi utama, yaitu dimensi fisik dan dimensi perilaku. Pada dimensi fisik, anomali mengacu pada penyimpangan atau perbedaan yang tidak lazim dari suatu bagian atau bahkan seluruh tubuh manusia sedangkan dari segi perilaku, anomali tersebut diadaptasi dari ilmu ekonomi, psikologi, dan juga ilmu sosiologi (Takahiro Nakao *et al.*, 2021).

2.3. Akurasi

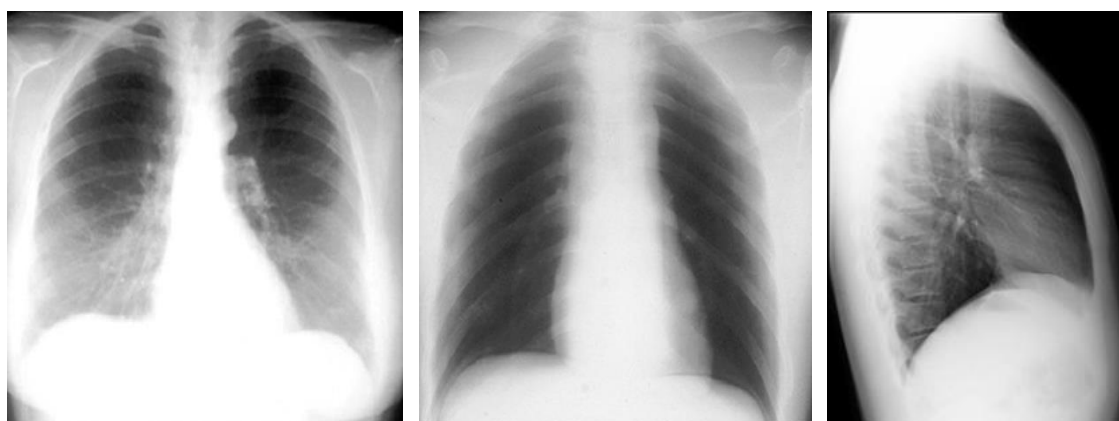
Akurasi adalah metrik untuk mengukur keberhasilan model klasifikasi. Akurasi juga merupakan pembagian prediksi model yang dianggap benar dengan jumlah total yang diprediksi. Perhitungan ini dapat digambarkan sebagai berikut (Klasifikasi: Akurasi 2018):

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Jumlah\ total\ prediksi}$$

2.4. Gambar Rontgen Dada

Rontgen dada merupakan metode pemeriksaan yang menggunakan sinar-X melalui alat *röntgen*. Metode ini menghasilkan citra dari organ-organ dan struktur di dalam dada, termasuk jantung, paru-paru, tulang, dan pembuluh darah. Tujuan dari Rontgen dada adalah untuk mengetahui kondisi organ, tulang, dan pembuluh darah di area dada, sehingga dokter dapat membantu mendiagnosis kondisi medis pasien. Terdapat beberapa jenis pemeriksaan rongga dada berdasarkan posisi tubuh. Untuk tampilan depan paru-paru, terdapat 2 jenis yaitu *Posterior-Anterior* (PA) dan *Anterior-Posterior* (AP). Sedangkan untuk tampilan samping, disebut juga sebagai *Lateral* (Candemir dan Antani, 2019).

Untuk memperoleh citra PA, seseorang akan berdiri dengan dada menempel pada mesin dan lengan direntangkan ke samping (Raoof et al., 2012). Sedangkan untuk memperoleh citra AP, seseorang harus berdiri atau terlentang di tempat tidur dengan punggung menempel pada mesin. Citra lateral digunakan sebagai pelengkap tampilan depan paru-paru karena citra lateral memberikan tampilan terbaik dari volume paru-paru secara keseluruhan (Raoof et al., 2012). Organ tubuh seperti tulang atau benda logam ditampilkan dengan warna putih dalam citra rontgen, sedangkan organ yang berisi ruang akan terlihat berwarna hitam. Contoh foto Rontgen dapat dilihat pada Gambar 2.1.



(a) PA

(b)AP

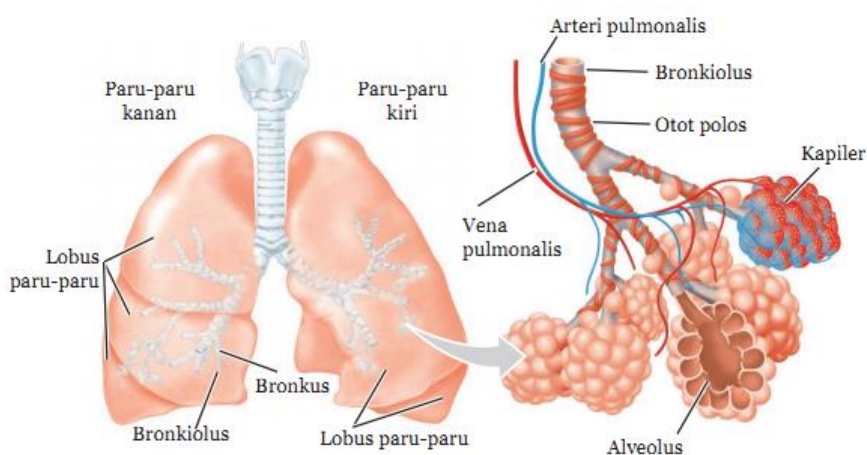
(c)Lateral

Gambar2.1 Rontgen Dada PA, AP, Lateral

(Sumber: <https://www.nejm.org>)

2.4. Paru-Paru

Paru-paru merupakan organ penting pada tubuh manusia yang memiliki fungsi sebagai sistem pernafasan (bronkus) yaitu sebagai tempat bertukarnya oksigen dengan kabondioksida di dalam darah. Apabila paru-paru mengalami masalah atau gangguan , maka akan mempengaruhi kesehatan pada tubuh dan organ lainnya pada manusia. Dalam tubuh manusia terdapat sepasang paru-paru yaitu paru-paru bagian kanan dan paru-paru bagian kiri yang mempunyai ciri berbeda, misalnya pada umumnya paru-paru bagian kiri orang dewasa memiliki bobot sekitar 325-550 gram, sedangkan paru bagian kanan mempunyai massa sekitar 375-600 gram. Lalu, pada paru-paru bagian kanan terdapat 3 bagian lobus (bagian yang menonjol dan bulat dari suatu organ), sedangkan pada paru-paru bagian kiri terdapat dua bagian lobus (Wira, 2021). Dapat dilihat pada gambar 2.2.



Gambar2.2 Pembagian lobus pada paru-paru

(Sumber: <https://www.bospedia.com>)

Paru-paru manusia berada di dalam rongga dada dan tergantung di antara tulang rusuk. Kedua paru-paru ini terletak di sisi-sisi jantung dan memiliki warna merah muda. Jika paru-paru telah rusak akibat udara yang telah tercemar pada akhirnya akan menimbulkan *patch* berbintik yang mewarnai paru-paru. Masing-masing paru-paru dilindungi oleh kantong yang disebut *pleura*. *Pleura* terdiri dari membran ganda yang terdiri dari membran serosa tipis. Bagian luar *pleura* disebut *pleura parietal* dan melekat pada dinding dada, sementara bagian dalamnya disebut *pleura visceral* dan meliputi paru-paru dan struktur terkait lainnya. Terdapat ruang di

antara dua membran tersebut yang disebut rongga *pleura*. (kelainan yang biasa ada pada paru-paru).

2.5. Kondisi Abnormal

Dalam penelitian ini terdapat beberapa kondisi abnormal pada paru-paru yang akan di klasifikasi, diantaranya :

2.3.1 Atelectasis

Atelectasis adalah penyakit paru dimana *alveolus* (struktur anatomi yang mempunyai bentuk berongga) yang tidak terisi oleh udara. Atau tempat terjadinya pertukaran oksigen dan karbondioksida sehingga apabila pertukaran oksigen dan karbon dioksida berjalan dengan baik, maka *alveolus* harus berisikan udara. *Atelectasis* juga dapat terjadi apabila kekurangan surfaktan pada dinding *alveolus*. Surfaktan merupakan zat yang berfungsi untuk *alveolus* tetap mengembang.

2.3.2 Cardiomegaly

Cardiomegaly adalah merupakan kondisi dimana pembesaran yang tidak biasa terhadap kondisi jantung yang tidak memiliki ukuran jantung pada umumnya. *Cardiomegaly* biasanya terjadi karena adanya pembesaran dari 4 ruangan jantung. Pada umumnya *cardiomegaly* diakibatkan karena adanya pembesaran yang terjadi pada bilik jantung sebelah kiri.

2.3.3 Consolidation

Consolidation merupakan kondisi yang terjadi apabila terdapat sebuah cairan yang dapat mengisi saluran udara kecil pada paru-paru. *Consolidation* terjadi ketika cairan mengisi saluran udara kecil paru-paru, menyebabkan sulit bernapas. Hal ini terjadi karena udara tidak dapat melewati *Consolidation*, yang menghalangi paru-paru untuk membawa dan mengeluarkan udara dari tubuh.

2.3.4 Mass

Mass adalah penyakit paru-paru yang didefinisikan sebagai bintik atau area *abnormal* dengan ukuran lebih besar dari 3 cm, sekitar 1,5 inci. Penyakit *mass* hampir sama dengan *nodule* hanya saja ukurannya yang membedakannya, dan penyakit *mass* memungkinkan memiliki kelainan yang lebih ganas dari pada *nodule*.

2.3.5 *Infiltration*

Infiltration merupakan zat yang lebih padat daripada udara, sama seperti darah, protein atau nanah, yang tertinggal di dalam parenkim (Jaringan Heterogen) paru-paru. *Infiltration* pada paru-paru biasanya sering berhubungan dengan penyakit *Pneumonia* dan *tuberculosis*. *Infiltration* adalah penyakit seperti nanah, darah, atau protein, seperti zat yang lebih padat daripada udara yang berada didalam paru.

2.3.6 *Nodule*

Nodule merupakan benjolan yang terbentuk di dalam paru-paru. Biasanya, *nodule* paru-paru ditemukan secara tidak sengaja melalui pemeriksaan radiologi seperti *Rontgen* dada atau tomografi komputer (CT scan). *Nodule* paru-paru dapat memiliki berbagai penyebab, baik yang bersifat jinak (non-kanker) maupun ganas (kanker).

2.3.7 *Pneumonia*

Pneumonia adalah suatu kondisi peradangan pada paru-paru yang disebabkan oleh infeksi. Gejala *pneumonia* bisa bervariasi, mulai dari ringan hingga berat. Beberapa gejala umum yang sering dialami oleh penderita *pneumonia* meliputi batuk berdahak, demam, dan kesulitan bernapas. *Pneumonia* juga dikenal dengan sebutan "paru-paru basah" karena infeksi menyebabkan peradangan pada *alveoli*, yaitu kantong udara di dalam paru-paru. Akibatnya, *alveoli* bisa terisi dengan cairan atau nanah, yang membuat penderita sulit bernapas (Cakti indra, 2020)

2.3.8 *Effusion*

Effusion diakibatkan oleh penumpukan cairan pada rongga pleuranya, pada rongga pleura terdapat lapisan yang menempel pada dinding yang berada pada rongga dada dan membungkus paru-paru.

2.3.9 *Pneumothorax*

Pneumothorax adalah kondisi yang terjadi apabila udara terkumpul pada rongga pleura, yaitu ruang yang terdapat diantara dinding dada dan juga paru-paru. Udara yang masuk diakibatkan karena adanya cedera yang terjadi pada dinding atau robekan jaringan paru-paru. Dampaknya dapat terjadi dan menyebabkan paru-paru mengempis dan tidak bias mengembang.

2.3.10 Fibrosis

Fibrosis terjadi ketika jaringan paru-paru rusak atau terluka. Jaringan yang kaku dan menebal dapat membuat lebih sulit untuk paru-paru bekerja. Seseorang yang mengalami *fibrosis* paru akan mengalami sesak napas yang lebih parah saat kondisinya memburuk.

2.3.11 Emphysema

Salah satu penderita penyakit paru obstruktif kronis atau (PPOK), emfisema menyebabkan gangguan persaluran pada pernapasan jangka panjang dan menyebabkan napas pendek yang sangat mengganggu karena faktor infeksi pada jaringan paru-paru. Penyakit ini disebabkan oleh kerusakan *alveoli*, kantung udara pada paru-paru, memiliki fungsi sebagai penampung untuk pertukaran udara oksigen dengan karbondioksida.

2.3.12 No_Finding

No_Finding adalah jenis *Rontgen* dada yang memiliki hasil *Rontgen* normal, tidak ada gejala penyakit apapun, sehingga yang dihasilkan bukanlah sebuah penyakit.

2.6. HOG (Histogram of Oriented Gradient)

Algoritma *Histogram of Oriented Gradients* HOG adalah teknik yang menghitung nilai gradien dan orientasi pada bagian lokal citra. Ini melakukannya dengan memecah gambar lengkap menjadi beberapa wilayah yang lebih kecil dan menghitung nilai dari bagian lokal tersebut. Untuk metode ini, dataset yang digunakan sebagai data latih sangat diperlukan. Histogram lokal dinormalisasi melalui pembagian dengan jumlah besarnya gradien melintasi wilayah gambar yang lebih besar, yang disebut blok (Suman Sedai et al, 2015).

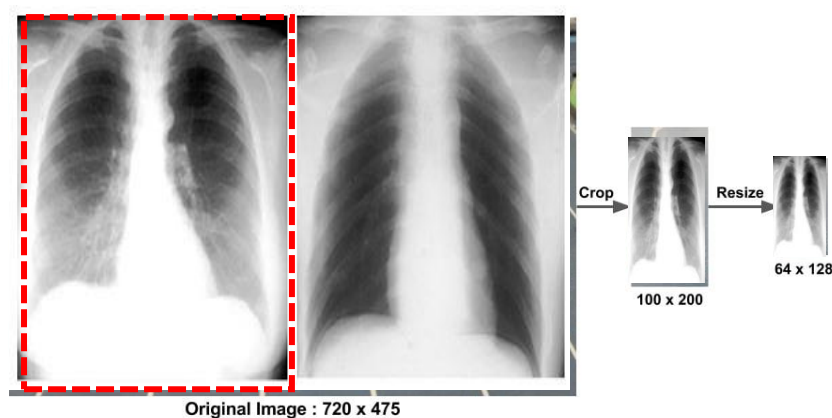
Deskriptor HOG berfokus pada struktur atau bentuk suatu objek. HOG tidak hanya klasifikasi tetapi juga mengekstraksi (memisahkan) arah tepi dengan menggunakan gradien dan orientasi dari tepi. Setiap wilayah gambar dipecah menjadi histogram yang dibuat berdasarkan nilai gradien dan orientasi piksel. Biasanya, deskriptor fitur mengubah gambar berukuran lebar kali tinggi x 3 (*channel*) menjadi vektor atau *array* fitur dengan panjang n . Dalam deskriptor fitur HOG, gambar

masukannya biasanya memiliki ukuran $64 \times 128 \times 3$ dan menghasilkan vektor fitur dengan panjang 3780, seperti yang dijelaskan dalam paper aslinya. Namun, deskriptor fitur HOG juga dapat dihitung untuk ukuran gambar yang lain. Berikut adalah langkah-langkah perhitungan untuk fitur HOG:

2.6.1 Preprocessing

Lakukan *Preprocessing* pada citra seperti konversi ke skala abu-abu, normalisasi intensitas, dan penghilangan *noise*. Hal ini dilakukan untuk mempersiapkan citra sebelum ekstraksi fitur. Langkah awal dalam perhitungan fitur HOG adalah memilih gambar input yang ingin dihitung fitur HOG-nya. Selanjutnya, gambar tersebut diubah ukurannya dengan rasio 1:2 (lebar:tinggi), sehingga memiliki ukuran 64×128 piksel. Ukuran ini dipilih karena memungkinkan pembagian gambar menjadi blok-blok dengan ukuran 8×8 dan 16×16 piksel saat mengekstrak fitur HOG. Penulis paper aslinya merekomendasikan dimensi ini sebagai target utama untuk mencapai hasil yang lebih baik dalam tugas klasifikasi pejalan kaki..

Untuk langkah pertama dapat dilihat pada gambar 2.3.



Gambar2.3 Klasifikasi pejalan kaki

(sumber: <https://learnopencv.com>)

2.6.2 Perhitungan Orientasi Gradien

Dalam perhitungan deskriptor HOG, terdapat dua metode yang dapat digunakan untuk menghitung orientasi arah horizontal dan vertikal. Metode pertama adalah

menggunakan operator sobel, sedangkan metode kedua adalah menggunakan *centered derivative*. (turunan terpusat) yang dapat dilihat pada gambar 2.4.

kernel operator centered derivative

$$dx = \begin{bmatrix} -1 & 0 & +1 \end{bmatrix} \quad dy = \begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix}$$

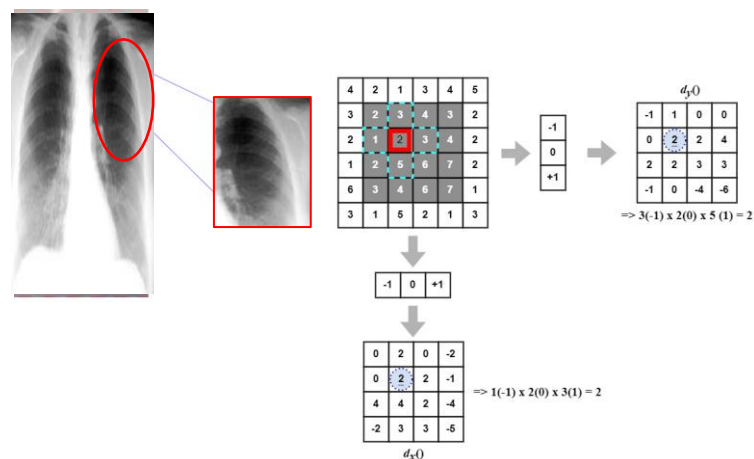
Kernel operator Sobel

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{dan} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Gambar2.4 rumus operator sobel atau *centered derivative*

(Sumber: <https://dosenit.com>)

Gambar berikut merupakan contoh perhitungan dalam menggunakan *centered derivative* (turunan terpusat). Proses konvolusi dapat dilakukan seperti pada gambar 2.5.



Gambar2.5 Perhitungan turunan terpusat

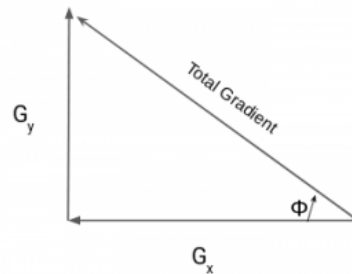
(sumber: <https://learnopencv.com>)

2.6.3 Menghitung *Magnitude* dan Gradien:

Selanjutnya mencari besaran dan arah gradien menggunakan rumus yang di buat pada persamaan 2.1 berikut.

$$\text{Magnitude } \mu = \sqrt{G_x^2 + G_y^2} \quad \text{Angle}(\theta) = |\tan^{-1}(G_y/G_x)| \quad (2.1)$$

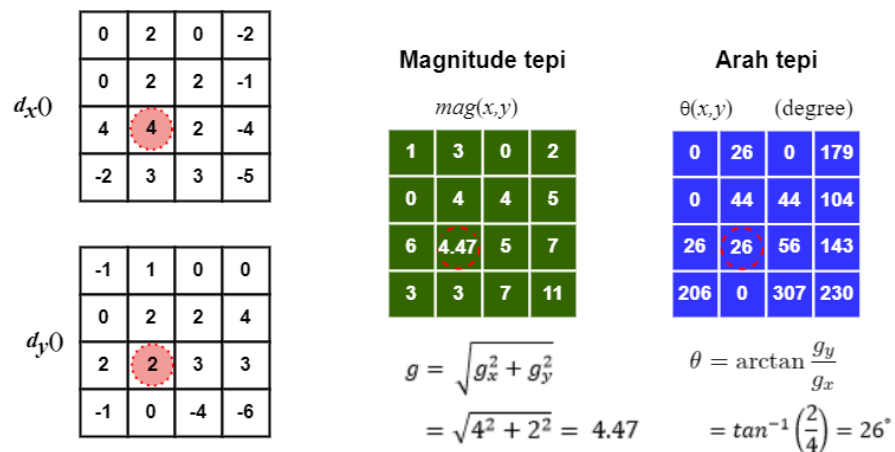
Dalam menghitung magnitude, digunakan *rumus teorema Pythagoras*.



Gambar2.6 Rumus Teorema Pitagoras

(Sumber: www.sefidian.com)

Setelah melakukan perhitungan berdasarkan orientasinya, langkah selanjutnya adalah menghitung nilai magnitude tepi (*edge magnitude*) dan arah tepi gradien (*edge direction*). Berikut ini contoh hasil perhitungan tersebut:

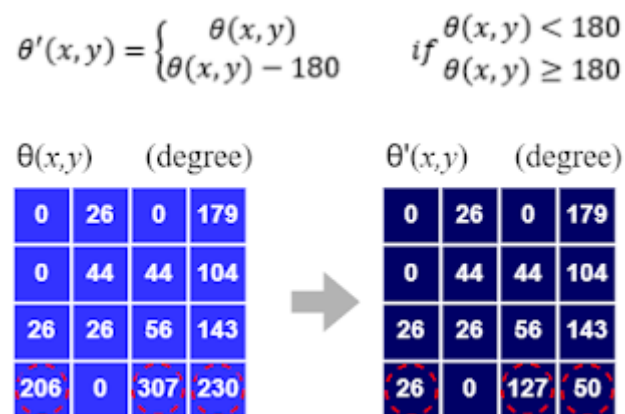


Gambar2.7 Perhitungan Nilai Magnitude

(Sumber: trivusi.web.id)

2.6.3 Membuat Histogram menggunakan Gradien dan Orientasi

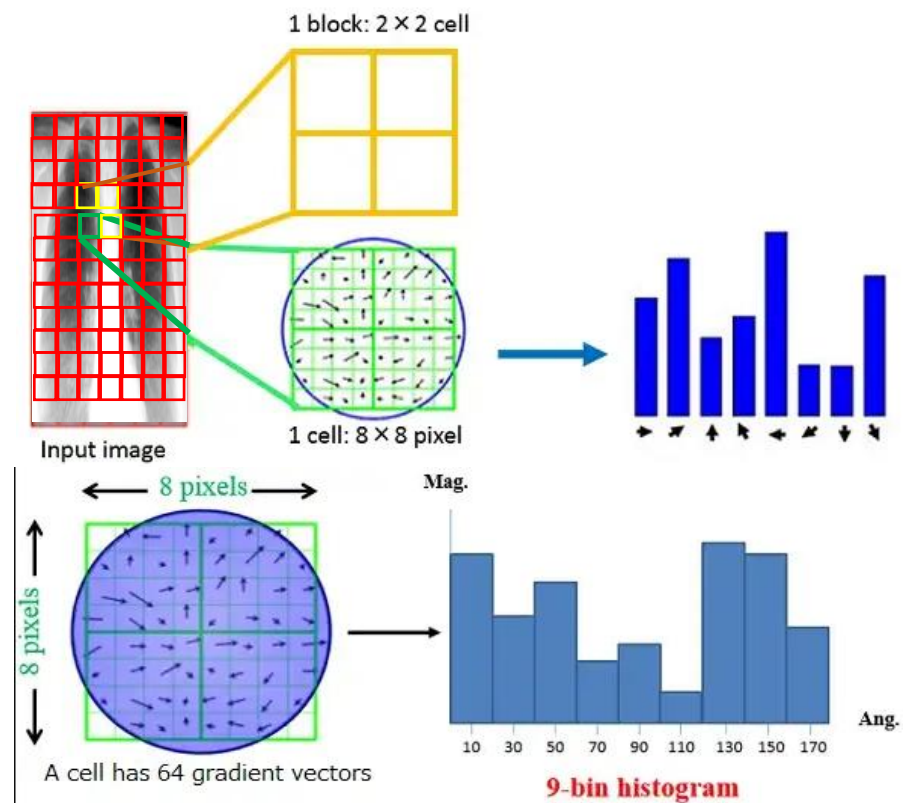
Untuk histogram, penting untuk membaginya menjadi 9 bin terpisah, mewakili sudut dari 0 hingga 160 derajat dengan kelipatan 20. Jika nilai pada setiap sel dari proses arah tepi melebihi atau sama dengan 180, maka dilakukan proses perubahan polaritas (*polarity change*). Perubahan polaritas tersebut bertujuan untuk membuat deskriptor menjadi invarian terhadap perubahan relatif antara *foreground* (objek) dan *background* (latar belakang)..



Gambar2.8 Proses Polarity change

(Sumber: *trivusi.web.id*)

Dalam deskriptor HOG, penting untuk membagi gambar menjadi sel-sel agar histogram gradien dapat dihitung untuk setiap sel. Sebagai contoh, jika gambar awalnya memiliki ukuran 64 x 128 piksel, maka gambar akan dibagi menjadi 8 x 8 sel. Perhatikan gambar 2.8 berikut ini!



Gambar2.9 Perhitungan cell HOG

(<https://www.trivusi.web.id>)

Terdapat 4 metode dalam pembuatan histogram diantaranya:

1. Metode 1: Dalam metode pertama, nilai dari setiap piksel akan diambil, orientasi piksel akan dihitung, dan tabel frekuensi akan diperbarui.

121	10	78	96	125
48	152	68	125	111
145	78	85	89	65
154	214	56	200	66
214	87	45	102	45

Frequency						1													
Angle	1	2	3	4 ...	35	36	37	38	39....					175	176	177	178	179	180

Gambar2.10 Metode Pembuatan Histogram 1

(<https://www.trivusi.web.id>)

Titik merah yang terlihat pada gambar menunjukkan sebuah piksel dengan *magnitude* 85 dan orientasi sudut 36. Oleh karena itu, pada kolom sudut 36 dari tabel frekuensi, nilai frekuensi akan diberikan sebagai 1. Proses ini akan diulangi untuk setiap piksel, dan pada akhirnya kita akan mendapatkan tabel frekuensi yang mencatat sudut dan jumlah kemunculannya dalam gambar. Tabel frekuensi ini dapat digunakan untuk membuat histogram dengan nilai sudut pada sumbu x dan frekuensi pada sumbu y . Perhatikan bahwa di sini nilai bin dari histogram adalah 1.

2. Metode 2: Metode ini mirip dengan metode sebelumnya, tetapi dengan ukuran bin yang merupakan kelipatan 20. Jadi, total bin yang digunakan adalah 9. Seperti sebelumnya, untuk setiap piksel, kita akan mengecek orientasinya atau arah tepi, dan menyimpan frekuensi nilai orientasi tersebut dalam bentuk matriks berukuran 9×1 .

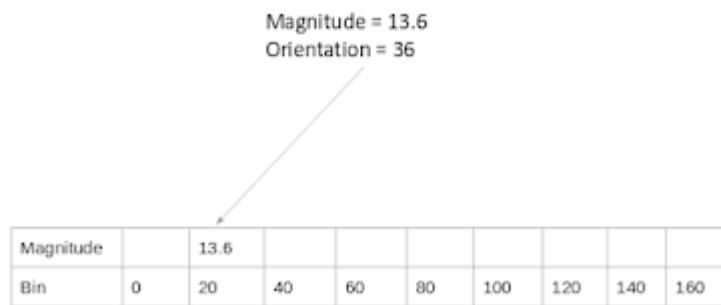
121	10	78	96	125
48	152	68	125	111
145	78	85	89	65
154	214	56	200	66
214	87	45	102	45

Magnitude		1							
Bin	0	20	40	60	80	100	120	140	160

Gambar2.11 Metode Pembuatan Histogram 2

(<https://www.trivusi.web.id>)

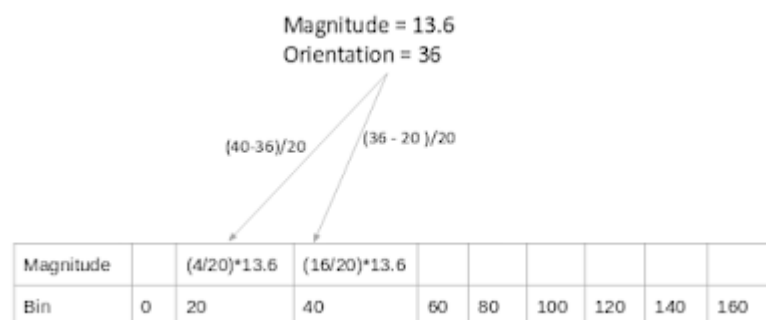
Metode 3: Kedua metode sebelumnya hanya mempertimbangkan nilai orientasi dalam pembentukan histogram, tanpa memperhitungkan nilai gradien. Namun, metode ketiga dapat menghasilkan histogram yang tidak hanya menggunakan frekuensi, tetapi juga memanfaatkan *magnitude* gradien untuk mengisi nilai dalam matriks.



Gambar2.12 Metode Pembuatan Histogram 3

(<https://www.trivusi.web.id>)

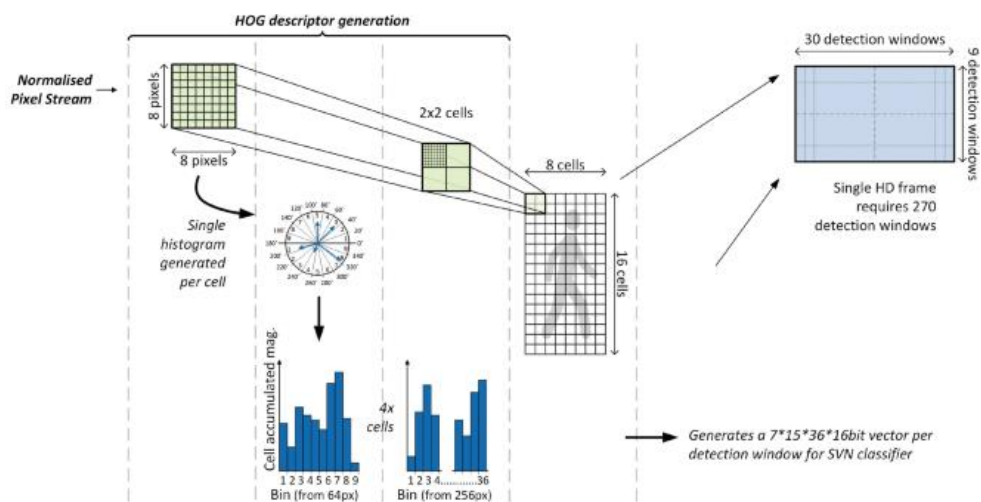
Metode keempat: Metode ini adalah salah satu metode yang paling umum digunakan dalam operasi HOG. Metode ini melakukan peningkatan dari metode ketiga sebelumnya. Penting untuk dicatat bahwa kontribusi yang lebih besar seharusnya diberikan pada bin yang lebih dekat dengan orientasi yang ada.



Gambar2.13 Metode Pembuatan Histogram 4

(<https://www.trivusi.web.id>)

Pada gambar tersebut, orientasi 36 berada di antara bin 20 dan 40. Dalam metode ini, dilakukan perhitungan nilai magnitudo yang akan dibagi dengan interval 20 (default) terhadap dua nilai bin yang berdekatan. Nilai magnitudo terbesar akan dikategorikan pada bin yang paling mendekati nilai orientasi tersebut. Contoh lain dari metode ini dapat dilihat pada gambar berikut:

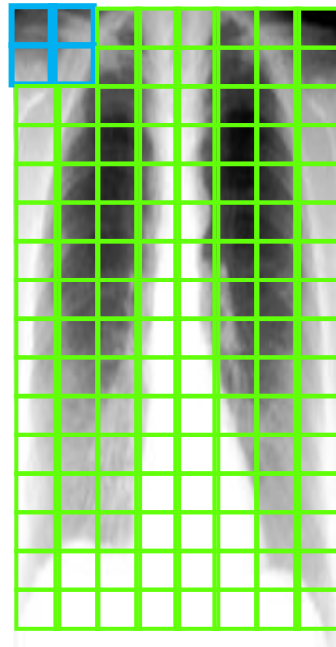


Gambar2.15 Perhitungan HOG 8 x 8 cell

(<https://www.trivusi.web.id>)

2.6.5 Normalisasi Gradien

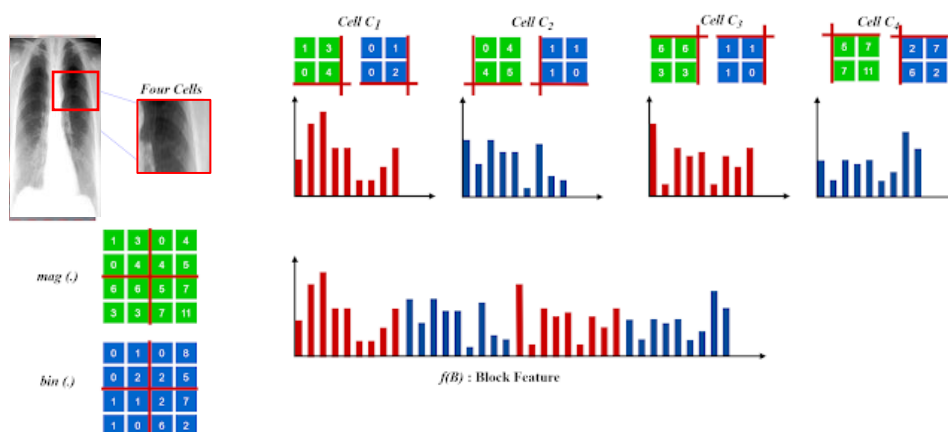
Meskipun fitur HOG telah dibuat untuk sel-sel gambar 8x8, terdapat situasi di mana perhitungan gradien menghasilkan nilai gradien yang bervariasi, dan gambar dapat menjadi sensitif terhadap pengaruh cahaya. Hal ini berarti bahwa dalam gambar tertentu, beberapa bagian gambar mungkin terlihat lebih terang dibandingkan dengan bagian lainnya. Meskipun tidak mungkin sepenuhnya menghilangkan variasi pencahayaan ini dari gambar, kita dapat mengurangnya dengan melakukan normalisasi gradien menggunakan blok 16x16.



Gambar2. 16 Normalisasi Gradien

(sumber: <https://learnopencv.com>)

Dalam proses ini, kita akan menggabungkan empat sel 8x8 untuk membentuk sebuah blok 16x16. Sebagai tambahan informasi, kita sudah mengetahui bahwa setiap sel 8x8 memiliki matriks histogram berukuran 9x1. Oleh karena itu, dengan menggabungkan empat sel tersebut, kita akan memiliki empat matriks berukuran 9x1, atau dalam kata lain, satu matriks berukuran 36x1.



Gambar2. 17 Menggabungkan 4 cell

(<https://www.trivusi.web.id>)

Untuk melakukan normalisasi pada matriks ini, kita akan membagi setiap nilai dalam matriks dengan akar kuadrat dari jumlah kuadrat pada nilai-nilai tersebut. Secara matematis, untuk sebuah vektor V yang diberikan, proses normalisasi akan dilakukan dengan membagi setiap elemen vektor tersebut dengan akar kuadrat dari jumlah kuadrat elemen-elemen vektor tersebut:

$$V = [a_1, a_2, a_3, \dots, a_{36}]$$

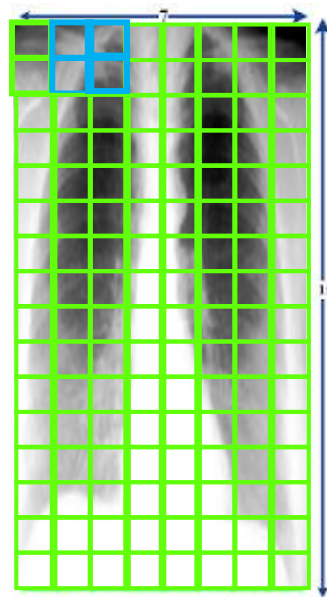
$$f(B) = \frac{2}{Z} \{v_0, v_1, \dots, v_{35}\} \text{ L1 - Norm: } Z = (|v_0| + |v_1| + \dots + |v_{35}|)$$

$$\text{Normalization Term L2 - Norm: } Z = \sqrt{(v_0)^2 + (v_1)^2 + \dots + (v_{35})^2} \quad (2.2)$$

Hasilnya akan menjadi vektor yang dinormalisasi dengan ukuran 36×1 .

2.6.6 Fitur untuk gambar lengkap

Ini merupakan tahap terakhir dalam pembentukan fitur HOG untuk gambar. Sampai saat ini, kita telah membuat fitur untuk blok gambar berukuran 16×16 . Sekarang, langkah selanjutnya adalah menggabungkan semua fitur ini untuk mendapatkan fitur akhir untuk gambar tersebut. Pertama-tama, kita perlu mengetahui berapa banyak blok 16×16 yang akan kita dapatkan untuk satu gambar berukuran 64×128 :

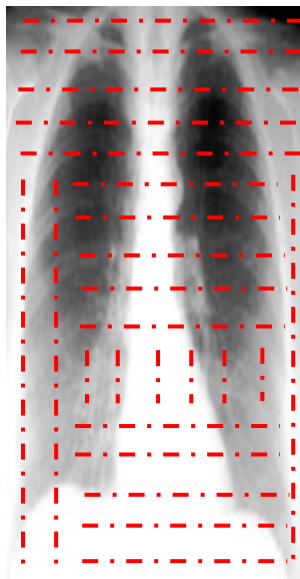


Gambar2. 18 Hasil Fitur HOG

(sumber: <https://learnopencv.com>)

Kita akan mendapatkan 105 (7 x 15) blok berukuran 16x16 dari gambar tersebut. Setiap blok akan memiliki vektor fitur berukuran 36x1. Oleh karena itu, total fitur untuk gambar tersebut akan menjadi $105 \times 36 \times 1 = 3780$ fitur.

Ini merupakan hasil dari proses deskriptor HOG. ditunjukkan pada gambar di bawah ini:



Gambar2. 19 Proses Deskriptor HOG

(sumber: <https://learnopencv.com>)

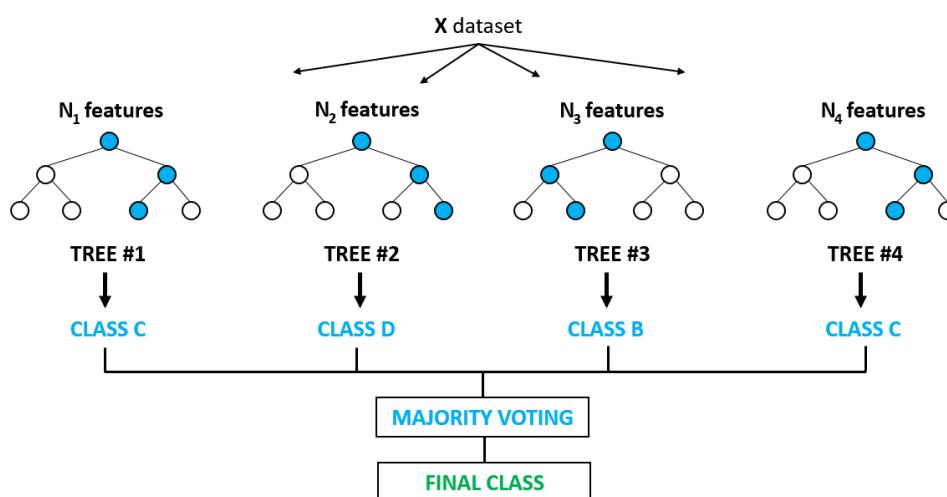
2.7 *Random forest*

Random Forest merupakan suatu algoritma yang berada didalam pembelajaran mesin yang digunakan untuk mengklasifikasikan dataset yang memiliki jumlah yang besar. Algoritma ini memiliki kemampuan untuk bekerja pada berbagai dimensi dan skala data, dan memiliki kinerja yang tinggi. Proses klasifikasi dilakukan dengan menggabungkan pohon-pohon dalam *Decision Tree* melalui pelatihan (*training*) dataset.

Salah satu algoritma terbaik dalam *machine learning* ini menggunakan *Decision Tree* untuk melangsungkan proses seleksi, di mana *Decision Tree* akan dibagi dengan proses pengulangan berdasarkan data dengan kelas yang sama. Dalam hal ini, penggunaan *tree* yang semakin banyak akan memiliki pengaruh terhadap akurasi yang didapat menjadi lebih optimal. Klasifikasi dengan menggunakan *Random Forest* dilakukan melalui hasil voting dari pohon-pohon yang

terbentuk. Hasil *voting* ini digunakan untuk menentukan klasifikasi akhir dari data yang diberikan.

Random forest merupakan algoritma *supervised learning* yang dipublikasikan oleh Leo Breiman (2001). Maksud dari *supervised learning* adalah sebuah tipe algoritma *machine learning* dengan menggunakan data berlabel untuk memberikan hasil prediksi. Algoritma *Random forest* merupakan tipe pembelajaran berdasarkan pada teknik dasar *data mining* yaitu algoritma *Decision Tree* (DT). *Random forest* dapat dilakukan secara *bagging* atau *aggregation bootstrap*, di mana proses tersebut berada di tengah-tengah proses pembagian data. *Bagging* merupakan teknik yang dibuat untuk meningkatkan stabilitas serta akurasi dari algoritma *Machine learning* dengan memanfaatkan *bootstrap sampling*. Peneliti bernama Breiman telah menggabungkan prinsip – prinsip dasar *bootstrap aggregating (bagging)* dengan pemilihan fitur acak untuk menambah keragaman (Q. Wang *et al*, 2019). *Workflow* dari algoritma *Random forest* terdapat pada Gambar 2.20 berikut ini.



Gambar2. 20 Workflow dari Algoritma *Random forest* (Yang *et al*, 2019)

(Sumber: <https://www.researchgate.net>)

Random Forest bekerja dengan membangun beberapa *Decision Tree* dan menggabungkannya untuk menghasilkan prediksi yang lebih stabil dan akurat. Seperti namanya, "Hutan" yang dibangun oleh *Random Forest* adalah kumpulan *Decision Tree* yang biasanya dilatih menggunakan metode *bagging*. Ide dasar dari metode

bagging adalah menggabungkan beberapa model pembelajaran untuk meningkatkan hasil keseluruhan.

Pada metode *bagging* akan menggabungkan seluruh eksperimen dengan jumlah suara terbanyak dan kelas yang paling banyak diprediksi. Namun pada saat membangun pohon identifikasi, proses pendekatannya adalah dengan membagi masalah menjadi sub masalah. Keputusan pada simbol teratas dibagi menjadi dua simbol, dan dua simbol tersebut terdiri dari pernyataan benar dan salah.

Algoritma *Random forest* meningkatkan keacakan pada model sambil menumbuhkan *tree*. Alih-alih mencari fitur yang paling penting saat memisahkan sebuah *node*, *Random forest* mencari fitur terbaik di antara *subset* fitur yang acak. Alhasil, cara ini menghasilkan keragaman yang luas dan umumnya menghasilkan model yang lebih baik. Tahapan pada pembangunan *Random forest* didalam kelompok data terdiri dari n pengamatan dan variabel prediktor (p):

1. Tahap *bootstrapping* : Mengambil D_i (dengan sampel berukuran n) secara *random* dari kumpulan data D dengan pengembalian menggunakan teknik *resampling bootstrap*.
2. Tahap *subsetting* secara random: membuat pohon berdasarkan data hasil *bootstrapping*, dengan *node* teratas ditentukan berdasarkan variabel prediktor (p) yang dipilih secara *random*. Jumlah variabel yang dipilih secara *random* dapat ditentukan dengan Persamaan 2.1 (Breiman, 2001), di mana M adalah jumlah variabel prediktor dan p adalah variabel prediktornya, dapat digunakan untuk menghitung jumlah variabel yang dipilih secara random.

$$M = \sqrt{p} \quad (2.3)$$

3. Tahap *predicting*: mengantisipasi sampel D_i berdasarkan pohon yang terbentuk pada langkah 2.
4. Mengulangi langkah 1 hingga 3 sebanyak k kali sampai Anda mendapatkan jumlah pohon yang diinginkan sebesar k .
5. Menggabungkan pohon-pohon yang dibuat pada langkah 4 untuk membuat prediksi dengan sampel data yang berbeda, dan kemudian membuat keputusan berdasarkan hasil vote terbanyak untuk output akhir untuk setiap pohon.
6. Hasil akhir dari algoritma *Random forest* adalah target kelas yang paling banyak diprediksi.

Setiap *decision tree* terdiri dari *internal node*, *leaf node* dan *root node*. *Internal node* yang merupakan *node* percabangan terdiri dari satu *input* dan minimal dua *output*, *leaf node* adalah *node* terakhir yang hanya memiliki satu *input*, dan *root Node* merupakan *node* teratas atau akar dari suatu pohon dan tidak mempunyai *output*. Pada setiap pohon keputusan setiap *leaf node* menandai label kelas. Algoritma *Random forest* akan menciptakan kumpulan pohon klasifikasi dengan ukuran dan bentuk yang berbeda. *Output* dari algoritma ini adalah kumpulan pohon klasifikasi yang mempunyai korelasi antar pohon yang kecil. Korelasi yang kecil akan mengurangi hasil kesalahan prediksi *Random forest*. Fitur lain dari *Random forest* adalah dapat menangani suatu data berdimensi tinggi, memiliki ketahanan untuk menghindari *overfitting* dan akan menggunakan pohon dalam jumlah yang besar selama proses kombinasi.

2.9. Hyperparameter Tuning

Hyperparameter tuning adalah proses mencari kombinasi optimal dari *hyperparameter* dalam suatu model atau algoritma *machine learning*. *Hyperparameter* adalah parameter yang tidak langsung dipelajari oleh model selama proses *training*, tetapi digunakan untuk mengatur perilaku dan kinerja model. Dalam *machine learning*, setiap algoritma memiliki *hyperparameter* yang dapat disesuaikan untuk mengoptimalkan performa model. Beberapa contoh *hyperparameter* yang umum digunakan termasuk *learning rate*, jumlah *hidden layer* dalam *neural network*, jumlah *estimator* dalam *Random forest*, dan lain sebagainya.

Tujuan dari *hyperparameter* tuning adalah mencari kombinasi *hyperparameter* yang memberikan hasil terbaik dalam hal akurasi, generalisasi, atau metrik evaluasi lainnya. Proses ini melibatkan eksplorasi dan evaluasi berbagai kombinasi *hyperparameter* untuk menemukan yang paling optimal. Ada beberapa metode yang dapat digunakan untuk melakukan *hyperparameter* tuning, antara lain:

- a) **Manual Tuning:** Melibatkan penyesuaian *hyperparameter* secara manual berdasarkan intuisi dan pengalaman praktisi *machine learning*. Namun, metode ini dapat memakan waktu dan tidak efisien karena mencoba banyak kombinasi *hyperparameter* secara manual.

- b) **Grid Search:** Menggunakan *grid search*, kita menentukan kumpulan nilai yang mungkin untuk setiap *hyperparameter*. *Grid search* akan mencoba semua kombinasi *hyperparameter* dalam grid dan menghitung performa model pada setiap kombinasi untuk memilih yang terbaik.
- c) **Random Search:** Seperti *grid search*, tetapi dalam random search, kita secara acak memilih kombinasi *hyperparameter* untuk dievaluasi. Ini mengurangi jumlah kombinasi yang harus dieksplorasi dibandingkan dengan *grid search*, tetapi masih mempertahankan kesempatan untuk menemukan kombinasi *hyperparameter* yang baik.
- d) **Bayesian Optimization:** Metode ini menggunakan model probabilitas untuk mencari kombinasi *hyperparameter* yang optimal. Model tersebut terus diperbarui berdasarkan informasi yang diperoleh dari setiap evaluasi *hyperparameter* yang dilakukan. Dengan demikian, metode ini mengarahkan pencarian ke wilayah yang paling menjanjikan.

Hyperparameter tuning penting untuk meningkatkan performa dan generalisasi model *machine learning*. Dengan memilih kombinasi *hyperparameter* yang optimal, kita dapat meningkatkan akurasi prediksi, mengurangi *overfitting*, dan mengoptimalkan kinerja model dalam berbagai tugas.

Model *hyperparameter* adalah pengaturan eksternal yang memiliki nilai tertentu, namun nilai-nilai tersebut tidak dapat ditentukan secara langsung dari data. Nilai-nilai *hyperparameter* model ditentukan sebelum proses pelatihan dimulai. Nilai-nilai ini tidak dapat ditemukan atau ditentukan dengan menganalisis data secara langsung. *Hyperparameter* dapat mempengaruhi kecepatan dan akurasi proses pembelajaran model, sehingga memilih nilai yang tepat untuk *hyperparameter* sangat penting dalam mengoptimalkan performa model. (Gravita, 2022).

2.8 Penelitian Terdahulu

Sebelumnya beberapa pakar penelitian telah membuat penelitian yang berkaitan dengan klasifikasi suatu penyakit. Berikut ini merupakan beberapa penelitian telah dilakukan untuk klasifikasi penyakit, klasifikasi objek-objek dalam dunia nyata dengan menggunakan algoritma *Random forest*. Penelitian untuk mengklasifikasi penyakit pada *Rontgen* dada sudah dilakukan sebelumnya baik menggunakan algoritma *Random forest* maupun algoritma lainnya:

Seperti yang sudah dilakukan pada penelitian (Saprudin *et al*, 2021) yang berkaitan dengan klasifikasi citra yang menggunakan metode *Random forest* dan sequential minimal optimization (SMO), Pada penelitian ini, metode klasifikasi dengan algoritma *Random forest* dan SMO tidak jauh beda dengan metode *Random forest* untuk scenario split 66% mendapatkan nilai akurasi sebesar 92.68%.

Pada penelitian lain dilakukan oleh (Mustopa *et al*, 2022) yang membahas mengenai penggunaan Ekstraksi Fitur dengan *Classifier Random forest* untuk memprediksi Covid-19 berdasarkan hasil rontgen thoraks dikategorikan sebagai penelitian yang baik. Hal ini dikarenakan penelitian ini mencapai akurasi tertinggi sebesar 84% dengan menggunakan algoritma klasifikasi *Random forest*, yang juga memiliki Standar Deviasi sebesar 0.015847. Selain itu, penelitian ini juga menghasilkan nilai Kappa sebesar 0.713. Proses mencapai akurasi tersebut melibatkan beberapa tahapan, yaitu menggunakan ekstraksi fitur seperti hu-moments, Harralick, dan histogram.

Selanjutnya penelitian yang dilakukan oleh (Ita Yuliant *et al*, 2022) Penelitian ini bertujuan untuk menganalisis efektivitas Algoritma *Random forest* dengan teknik SMOTE dalam memprediksi risiko kanker paru-paru. Dalam penelitian ini, pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman *Python*. Hasil pengujian menunjukkan tingkat akurasi sebesar 88% dengan nilai AUC (*Area Under Curve*) sebesar 0,93. Saat menggunakan algoritma *Random forest* untuk memprediksi risiko kanker paru-paru, teknik SMOTE terbukti efektif dalam mengatasi ketidakseimbangan kelas dalam set data yang digunakan.

Selanjutnya penelitian yang dilakukan (Nanik Wuryani *et al*, 2021) Penelitian ini fokus pada penggunaan *Random forest Classifier* untuk mengklasifikasi penderita COVID-19 berdasarkan citra CT Scan. Penelitian ini dimulai dengan menguji satu fitur secara terpisah, kemudian mengombinasikannya dengan fitur lain, dan membandingkannya dengan klasifikasi menggunakan algoritma lain seperti *K-Nearest Neighbor* (KNN), *Decision Tree*, *Linear Discriminant Analysis* (LDA), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Naive Bayes*.

Hasil penelitian menunjukkan bahwa algoritma *Random forest* paling sesuai untuk klasifikasi citra CT Scan paru-paru dengan memasukkan fitur *Harralick* dan *Color Histogram* ke dalam proses. Algoritma ini mencapai tingkat akurasi sebesar 96,9%, diikuti oleh KNN dengan tingkat akurasi sebesar 96,5%, *Decision Tree* dengan

tingkat akurasi sebesar 95,5%, dan *Naive Bayes* dengan tingkat akurasi terendah sebesar 42,4%.

Selanjutnya penelitian yang dilakukan (Alfiana Ramdhan *et al*, 2014) Penelitian ini membahas tentang klasifikasi citra rontgen dada dengan menggunakan metode ekstraksi fitur Histogram dan *Naïve Bayes Classifier*. Hasil pengujian menunjukkan bahwa metode *Naive Bayes Classifier* mampu mengenali pola masukan citra dan mengklasifikasikannya. Dibandingkan dengan penelitian sebelumnya yang menggunakan jaringan syaraf tiruan *backpropagation*, penelitian ini mencatat peningkatan akurasi sebesar 73%.

Selanjutnya penelitian yang dilakukan (Rahmat Hidayat *et al*, 2021), Penelitian ini mencakup penerapan metode pembelajaran untuk mengidentifikasi dan mengenali iris menggunakan ekstraksi fitur dan algoritma klasifikasi. Hasil penelitian menunjukkan bahwa nilai akurasi tertinggi diperoleh dengan menggabungkan ekstraksi fitur HOG dan *haralick* pada algoritma *Random forest*, dengan akurasi mencapai 81,38%.

Tabel 2. 1 Peneliti Terdahulu

No.	Penelitian	Judul	Metode	Keterangan
1.	Saprudin <i>et al</i> (2021)	Klasifikasi Citra Menggunakan Metode <i>Random forest</i> dan Sequential Minimal Optimization (SMO)	<i>Random forest</i> dan Sequential Minimal Optimization (SMO)	Pada penelitian ini, metode klasifikasi dengan algoritma <i>Random forest</i> dan SMO tidak berbeda jauh dengan metode <i>Random forest</i> untuk <i>scenario split</i> 66% mendapatkan nilai akurasi sebesar 92.68%
2.	Ali Mustopa <i>et al</i> (2022)	Ekstraksi Fitur dengan Classifier <i>Random forest</i> untuk Memprediksi Covid 19 Berdasarkan Hasil	<i>Random forest</i>	Penelitian ini termasuk kedalam kategori baik karena menghasilkan akurasi sebesar 84%. Hasil akurasi tersebut dilakukan dengan beberapa tahapan

		<i>Rontgen Thorax</i>		yaitu dengan <i>feature extraction</i> berupa <i>hu-moments</i> , <i>Harrallic</i> dan histogram
3.	Ita Yulianti <i>et al</i> (2022)	The Effectiveness Analysis of <i>Random forest</i> Algorithms With SMOTE Techique in predicting Lung Cancer Risk	<i>Random forest</i>	Pada penelitian ini pengolahan dan analisis data menggunakan bahasa pemrograman <i>Python</i> . Hasil pengujian menunjukkan nilai akurasi sebesar 88%.
4.	Nanik Wuryani <i>et al</i> (2021)	<i>Random forest</i> <i>Classifier</i> untuk Klasifikasi Penderita COVID- 19 berbasis Citra CT <i>Scan</i>	<i>Random forest</i>	Penelitian dimulai dengan memasukkan satu fitur ke dalam percobaan, lalu mengkombinasikan dengan fitur yang lain, kemudian membandingkannya menggunakan klasifikasi oleh algoritma lain seperti <i>K-Nearest Neighbor</i> (KNN), <i>Decision Tree</i> , <i>Linear Discriminant Analysis</i> (LDA), <i>Logistic Regression</i> , <i>Support Vector Machine</i> (SVM), dan <i>Naive Bayes</i> . Hasil penelitian menunjukkan bahwa algoritma <i>Random forest</i> paling cocok jika digunakan untuk klasifikasi citra CT <i>Scan</i> paru-paru dengan

				memasukkan fitur <i>Haralick</i> dan <i>Color Histogram</i> ke dalam proses yang menghasilkan akurasi sebesar 96,9%, diikuti oleh KNN sebesar 96,5%, <i>Decision Tree</i> sebesar 95,5%, dan yang paling rendah yaitu <i>Naive Bayes</i> sebesar 42,4%
5.	Alfiana Ramdhan et al (2014)	Klasifikasi Citra <i>Rontgen</i> dada dengan Ekstraksi Fitur Histogram dan Metode <i>Naive Bayes Classifier</i>	<i>Histogram</i> dan <i>Naive Bayes Classifier</i>	Pada proses pengujian didapatkan bahwa metode <i>Naive Bayes Classifier</i> dapat mengenali pola masukan citra dan dapat mengklasifikasikan citra. Jika dibandingkan dengan hasil penelitian sebelumnya menggunakan jaringan syaraf tiruan <i>backpropagation</i> , pada penelitian ini terjadi peningkatan akurasi yakni sebesar 73%
6.	Rahmat Hidayat et al (2021)	Penerapan Metode Pembelajaran Menggunakan Ekstraksi Fitur dan Algoritma Klasifikasi untuk Identifikasi Pengenalan Iris	HOG dan <i>Haralick</i>	Nilai akurasi terbaik diperoleh dari penggabungan ekstraksi fitur HOG dan <i>haralick</i> pada algoritma <i>Random forest</i> dengan nilai akurasi sebesar 81.38%

Metode yang akan digunakan dalam penelitian ini membedakan dari penelitian sebelumnya, yang disajikan dalam tabel 2.1. Pada tahap praproses, penelitian ini akan menggunakan dua metode: HOG untuk menghilangkan data *noise* dan *Random Undersampling* untuk menyeimbangkan data hingga mencapai titik rasio yang optimal. Karena data yang dihasilkan akan lebih baik dengan dua metode praproses yang digunakan. Peneliti kemudian akan menggunakan *Random forest* sebagai opsi terbaik untuk mengklasifikasikan data, seperti yang dilakukan pada penelitian sebelumnya.

BAB 3

ANALISIS DAN PERANCANGAN

3.1. Dataset yang digunakan

Data yang digunakan dalam penelitian ini adalah data *Rontgen* dada, data ini telah dicari oleh peneliti dan didapatkan dari situs yang disediakan dari kompetisi Kaggle.com. Data yang digunakan dalam kompetisi ini dibuat dengan mengumpulkan studi sinar-X Dada yang tidak teridentifikasi yang disediakan oleh dua rumah sakit di Vietnam: Rumah Sakit 108 dan Rumah Sakit Universitas Kedokteran Hanoi. Adapun total dari keseluruhan data yaitu 2.400 gambar *postero-anterior* (PA), yang mana total data telah difilter atau dipilah oleh peneliti.

Selanjutnya peneliti akan membagi menjadi dua bagian data yaitu data *training*, dan *testing*. Untuk data *training* sendiri diterapkan dalam proses pembelajaran algoritmanya dalam mengklasifikasi jenis penyakit paru. Sementara data *testing* digunakan untuk menguji algoritmanya. Adapun untuk pembagian data *training* sebesar 80% atau 1.920 data dan data *testing* sebesar 20% atau 480 data, seluruh data berekstensi PNG.


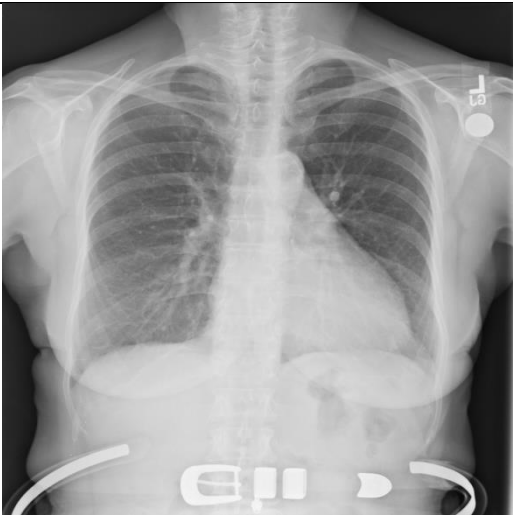
Tabel 3. 1 Jumlah Data


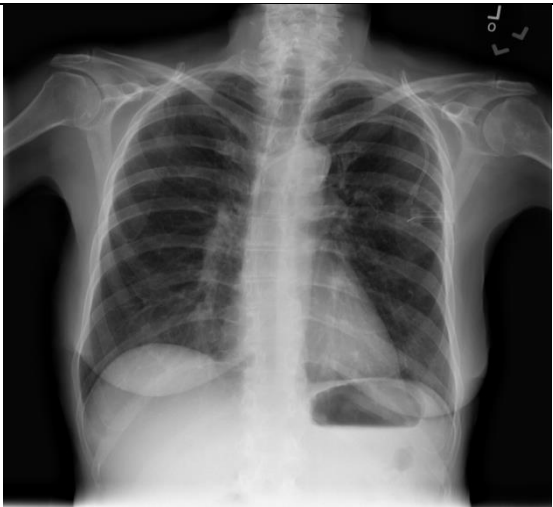

NO.	Jenis Objek Abnormal	Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>
1.	<i>Atelectasis</i>	160	40
2.	<i>Cardiomegaly</i>	160	40
3.	<i>Concolidation</i>	160	40
4.	<i>Infiltration</i>	160	40
5.	<i>Nodule</i>	160	40
6.	<i>Mass</i>	160	40
7.	<i>Emphysema</i>	160	40
8.	<i>Fibrosis</i>	160	40
9.	<i>Pleural effusion</i>	160	40
10.	<i>Pneumothorax</i>	160	40

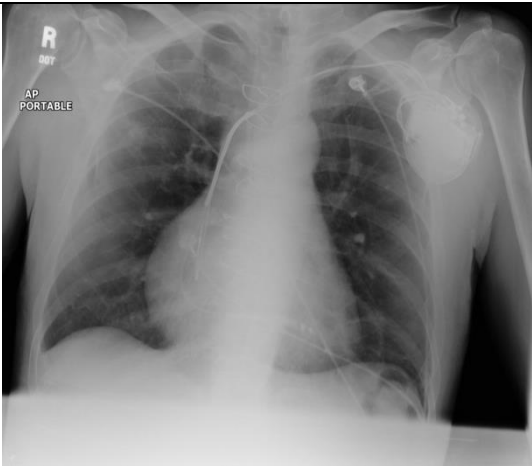
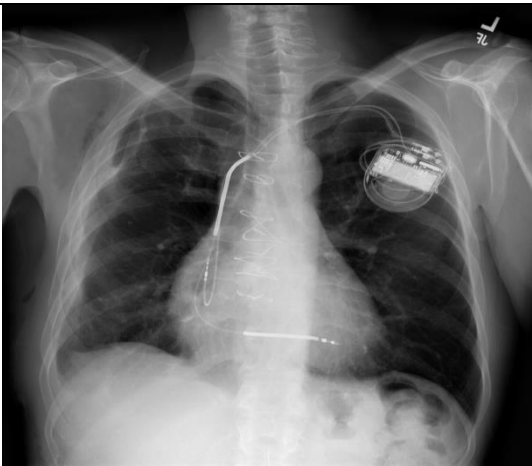

11.	<i>Pneumonia</i>	160	40
12.	<i>No_Finding</i>	160	40
Total Data		1.920	480


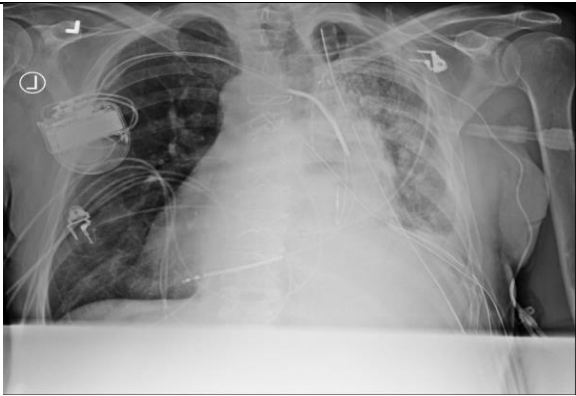
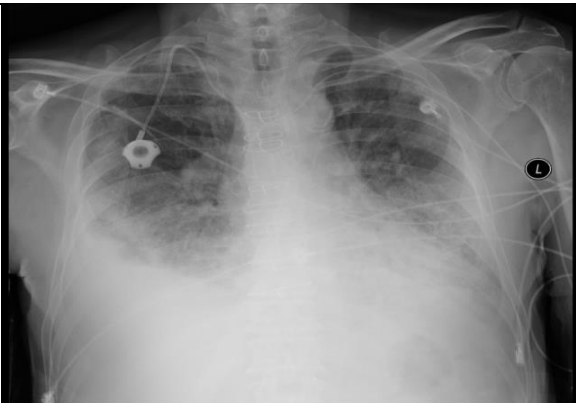
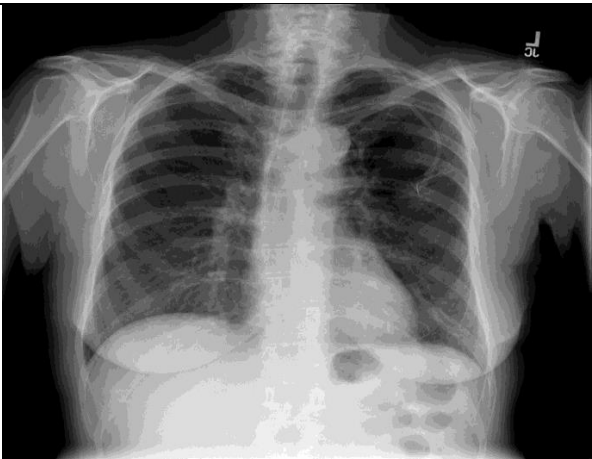
Dari tabel 3.1 jumlah keseluruhan data *training* adalah 1.920 dan data *testing* adalah 480 sehingga total keseluruhan data adalah 2.400.

Tabel 3. 2 Contoh Jenis hasil Radiologi Paru

NO.	Gambar Rontgen	Jenis Objek Abnormal
1.		<i>Atelectasis</i>
2.		<i>Cardiomegaly</i>

NO.	Gambar Rontgen	Jenis Objek Abnormal
3.		<i>Concolidation</i>
4.		<i>Infiltration</i>
5.		<i>Nodule</i>

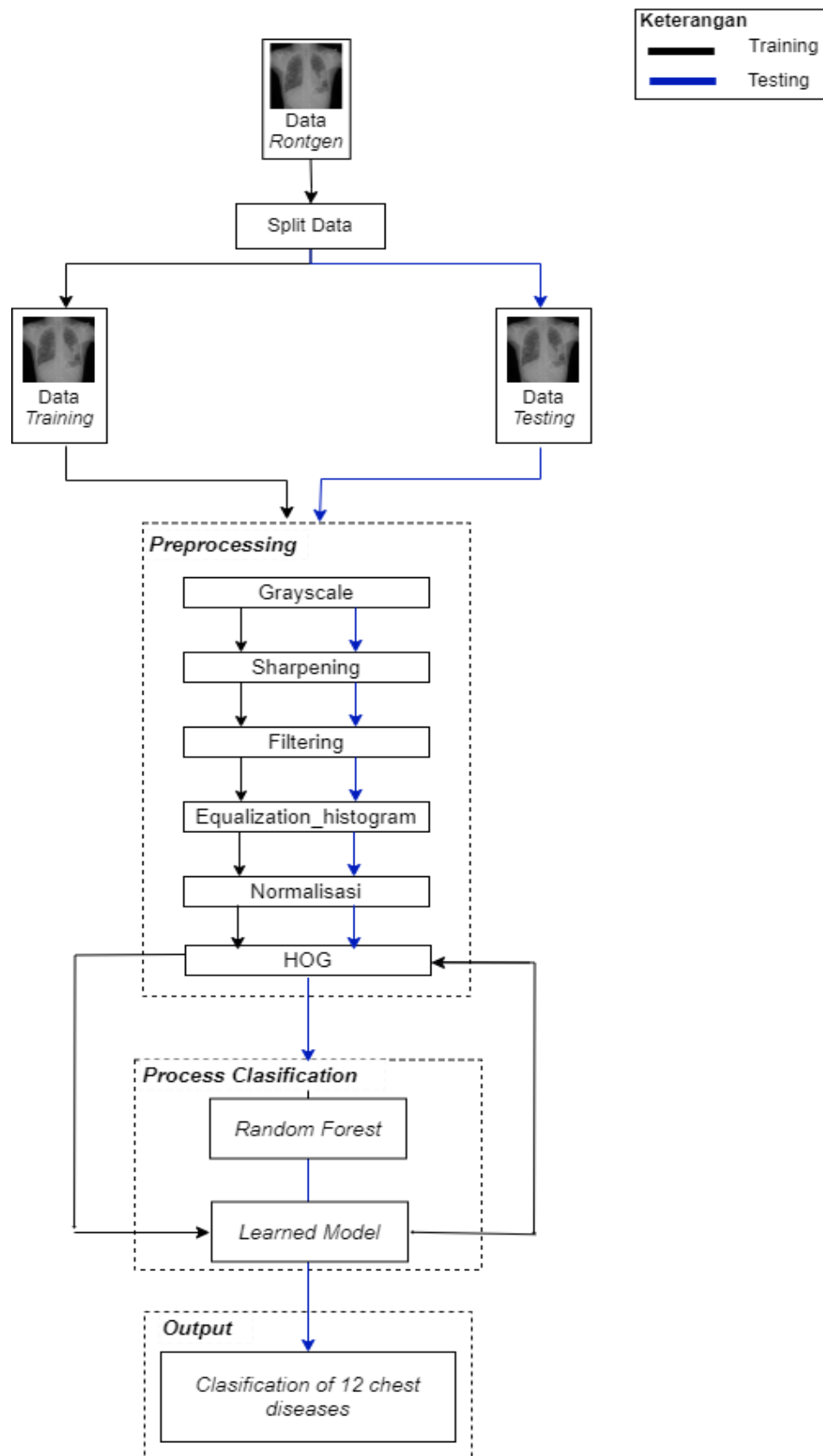
NO.	Gambar Rontgen	Jenis Objek Abnormal
6.		<i>Mass</i>
7.		<i>Emphysema</i>
8.		<i>Fibrosis</i>

NO.	Gambar Rontgen	Jenis Objek Abnormal
9.		<i>Effusion</i>
10.		<i>Pneumothorax</i>
11.		<i>Pneumonia</i>
12.		<i>No_Finding</i>

3.2. Arsitektur Umum

Dalam penelitian ini telah dilakukan beberapa langkah-langkah dalam kasus mengklasifikasi objek abnormal pada paru-paru. Langkah pertama yang telah dilakukan adalah pengumpulan data dalam bentuk citra digital DICOM atau (*Digital Imaging and Communications in Medicine*). Dalam pre-processing citra DICOM yang telah dikumpulkan akan dikonversi kedalam bentuk PNG (*Portable Network Graphics*) agar data dapat diproses. Selanjutnya dilakukan pre-processing terhadap dataset, seluruh data *training* yang berada dalam 1 folder di input untuk mengetahui hasil data citra asli, *grayscale*, *sharpening*, *filtering*, *equalizaion_histogram*, dan HOG.

Pada tahapan ini akan membahas bagian yang dilakukan dalam pembangunan suatu sistem untuk mengklasifikasi Radiologi pada gambar *Rontgen* dada. Adapun tahapan tersebut meliputi *data training* dan *data testing* menggunakan *Random forest* dan *output* nama penyakit dari data yang telah di input. Tahapan-tahapan tersebut telah detail dijelaskan pada arsitektur umum yang ada pada gambar 3.1.



Gambar3.1 Arsitektur Umum

Dalam penelitian ini terdapat beberapa tahapan yang akan dilakukan proses perancangan, yaitu sebagai berikut:

1. *Data Training*

Saat proses *training* selesai dilakukan dengan menggunakan arsitektur algoritma *Random forest*, maka hasil yang diharapkan adalah pembentukan model dari proses *training*. Selain membentuk model *Random forest*, hasil *training* adalah tingkat akurasi dari model tersebut.

2. *Data Testing*

Data *testing* adalah data yang digunakan untuk mengetes dan melihat keakuratannya, atau melihat performa data yang sebelumnya telah diuji.

3. *Preprocessing*

Preprocessing merupakan tahapan yang penting sebelum melakukan *training*. *Preprocessing* mempermudah data untuk diproses dan meningkatkan akurasi dari hasil *training*. Beberapa tahapan pada *pre-processing* yaitu :

- a. *Grayscale*

Mengacu pada representasi citra yang hanya memiliki tingkat keabuan (nilai intensitas) tanpa informasi warna. Citra *grayscale* dihasilkan dengan mengonversi citra berwarna menjadi citra dengan satu saluran intensitas saja.

- b. *Sharpening*

Proses untuk meningkatkan kejelasan dan kontras detail pada gambar. Tujuannya adalah untuk meningkatkan tampilan gambar dengan meningkatkan batas-batas dan tekstur yang terlihat dan untuk mempertajam *pixel*, karena semakin besar angka *pixel* maka data akan semakin akurat.

- c. *Filtering*

Memisahkan antara gambar dan objek latar belakang, dengan cara memisahkan warna yang paling terang dan merupakan salah satu teknik penting dalam pemrosesan citra yang digunakan untuk mengubah karakteristik spasial atau frekuensi dari sebuah citra. *Filtering* pada pemrosesan citra melibatkan penerapan operasi filter pada setiap piksel dalam citra untuk mencapai tujuan tertentu, seperti penghilangan derau, peningkatan ketajaman, atau ekstraksi fitur.

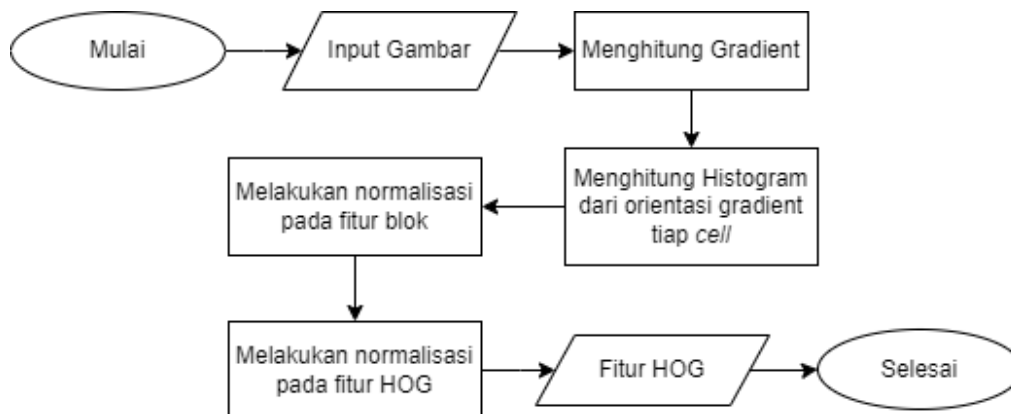
d. *Equalization_histogram*

Metode perbaikan kualitas citra untuk meratakan nilai intensitas pixel pada suatu citra dan merupakan teknik dalam pemrosesan citra yang digunakan untuk meningkatkan kontras dan distribusi intensitas piksel dalam citra. Tujuannya adalah untuk mengubah histogram citra menjadi distribusi intensitas yang lebih merata.

e. *HOG (Histogram of Oriented Gradient)*

Histogram of Oriented Gradients (HOG), merupakan metode untuk mengklasifikasi suatu objek dengan representasi bentuk. HOG bekerja dengan menangkap histogram dari orientasi gradien pada setiap *cell* yang terbagi dari sebuah citra (Yohannes et al., 2020). Dari berbagai penelitian menunjukkan bahwa metode HOG cocok dijadikan deskriptor untuk pencarian gambar. Metode HOG banyak digunakan pada computer vision. HOG adalah deskriptor berbasis window yang mendeteksi pada titik interest. Metode ini menghitung nilai gradien dalam daerah tertentu pada suatu citra. Setiap citra memiliki karakteristik yang ditunjukkan oleh distribusi gradien yang diperoleh dengan membagi citra ke dalam daerah kecil yang disebut *cell*. Tiap *cell* disusun dari sebuah histogram dari sebuah gradien. Kombinasi dari histogram ini dijadikan sebagai deskriptor yang mewakili sebuah obyek. Diagram alir algoritma HOG dapat dilihat pada Gambar 3.2

1. Tahap awal dari metode HOG adalah menghitung nilai gradien citra untuk membagi beberapa citra menjadi beberapa bagian.
2. Pada citra seperti konversi ke skala abu-abu, normalisasi intensitas, dan penghilangan *noise*. Hal ini dilakukan untuk mempersiapkan citra sebelum ekstraksi fitur
3. Pembentukan fitur HOG untuk gambar dan selanjutnya adalah menggabungkan semua fitur ini untuk mendapatkan fitur akhir untuk gambar tersebut.



Gambar 3.2 Alur kerja HOG

4. Process

a. *Random forest*

Random forest Classifier (RF) merupakan salah satu metode pohon gabungan yang digunakan untuk klasifikasi dan regresi. Pada metode *Random forest*, dilakukan pembangunan sejumlah pohon keputusan menggunakan data pelatihan dan keluaran kelas sebagai dasar. Setiap pohon dalam *Random forest* merupakan model untuk klasifikasi atau prediksi, tergantung pada tujuan yang diinginkan, baik itu klasifikasi untuk memprediksi kelas atau regresi untuk memprediksi nilai. (Muslikh et al., 2019). Berikut merupakan langkah kerja *random forest: Learned Model*

Merujuk pada model atau algoritma yang telah dilatih using data latih untuk mempelajari hubungan antara fitur citra dan tugas yang ingin diselesaikan. Tugas *learned model* seperti pengenalan objek, segmentasi, klasifikasi, dan pemrosesan lainnya adalah beberapa contoh tugas yang ingin diselesaikan. Model yang telah dilatih dapat digunakan untuk memprediksi atau melakukan tugas serupa pada gambar yang belum pernah dilihat sebelumnya.

5. Output

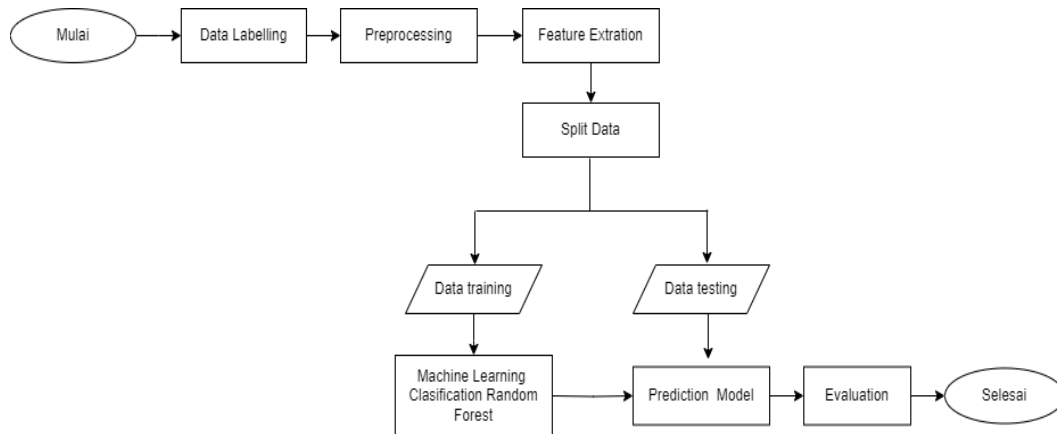
Menentukan hasil *Clasification of 12 critical radiographic* dan hasil akurasi. Proses "klasifikasi" dalam konteks ini merujuk pada pengelompokkan atau penentuan kategori dari *12 critical radiographic* yang telah diklasifikasikan. Selanjutnya, hasil akurasi akan mencerminkan sejauh mana klasifikasi yang dilakukan oleh model atau algoritma *random forest* cocok dengan klasifikasi yang sebenarnya dari hasil radiografi gambar yang sebenarnya.

3.3. Proses Training Random forest

Proses *training* pada *Random forest* melibatkan beberapa langkah utama. Berikut adalah langkah-langkah umum dalam proses *training Random forest*:

1. Persiapan data: Pertama-tama, data harus disiapkan dalam format yang sesuai untuk *Random forest*. Ini melibatkan pemisahan dataset menjadi fitur (*features*) dan target (target), serta pembagian data menjadi set pelatihan (*training set*) dan set validasi (*validation set*) jika diperlukan.
2. Pemilihan jumlah pohon: *Random forest* terdiri dari sejumlah pohon keputusan (*decision trees*). Pemilihan jumlah pohon adalah langkah penting dalam proses *training*. Jumlah pohon harus ditentukan sebelumnya dan dapat mempengaruhi kinerja model. Jumlah pohon yang lebih besar cenderung menghasilkan model yang lebih kuat tetapi juga membutuhkan lebih banyak waktu untuk melatih dan memprediksi.
3. Pelatihan pohon keputusan: Setiap pohon dalam *Random forest* dilatih secara independen menggunakan teknik *bootstrap aggregating (bagging)*. Ini melibatkan pemilihan sampel acak dengan penggantian dari set pelatihan untuk setiap pohon. Selanjutnya, setiap pohon dilatih menggunakan algoritma pembentukan pohon keputusan, seperti CART (*Classification and Regression Trees*) atau ID3 (*Iterative Dichotomiser 3*).
4. Pembentukan keputusan *ensemble*: Setelah semua pohon dilatih, hasil prediksi dari setiap pohon digabungkan untuk menghasilkan prediksi final. Untuk masalah klasifikasi, mode (modus) dari prediksi kelas dari semua pohon dapat diambil sebagai prediksi *ensemble*. Untuk masalah regresi, prediksi *ensemble* biasanya diambil sebagai rata-rata dari prediksi semua pohon.
5. Evaluasi dan tuning: Setelah pembentukan *ensemble*, model *Random forest* dievaluasi menggunakan metrik evaluasi yang sesuai, seperti akurasi (untuk klasifikasi) atau MSE (*Mean Squared Error*) (untuk regresi). Jika model tidak memenuhi kriteria kinerja yang diharapkan, beberapa parameter dapat dituning untuk meningkatkan performa, seperti jumlah fitur yang dipertimbangkan pada setiap pemisahan dalam pohon atau kedalaman maksimum pohon.

6. Penggunaan model: Setelah model *Random forest* dilatih dengan baik, ia dapat digunakan untuk melakukan prediksi pada data baru yang tidak pernah dilihat sebelumnya. Berikut merupakan langkah kerja *random forest*:



Gambar 3.3 Alur kerja *random forest*

3.4. Metode Evaluasi

Pada tahap ini, metode pengukuran performa atau metode evaluasi mempunyai peranan yang sangat penting dalam mengevaluasi kinerja suatu model *machine learning*. Pada umumnya, performa suatu model identifikasi dievaluasi dengan menghitung akurasi nya. Akurasi adalah ketepatan model dalam mengidentifikasi data dengan menghitung rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Berikut ini terdapat *confusion matrix* yang digunakan untuk mengukur kinerja model dalam memberi keputusan yang diperoleh pada proses *training* dan *testing*. Berikut adalah tabel *confusion matrix* seperti pada Tabel 3.6.

Tabel 3. 3 Confusion Matrix

		Nilai Prediksi	
		True	False
Nilai Aktual	True	TN (True Negative)	FP (False Positive)
	False	FN (False Negative)	TP (True Positive)

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100$$

dimana:

TP : Jumlah objek positif yang tepat diprediksi

TN : Jumlah objek negatif yang tepat diprediksi

FP : Jumlah objek positif yang salah diprediksi

FN : Jumlah objek negatif yang salah diprediksi

TP (True Positive) adalah jumlah kasus positif yang diklasifikasikan dengan benar. FN (False Negative) adalah jumlah positif kasus salah diklasifikasikan sebagai negatif. FP (False Positive) adalah jumlah kasus negatif yang salah diidentifikasi sebagai kasus positif dan TN (True Negative) adalah jumlah kasus negatif dengan benar diklasifikasikan. Dengan konvensi, kami menganggap kelas minoritas dalam pemodelan data tidak seimbang sebagai kelas positif sementara kelas mayoritas dianggap sebagai kelas negatif. Sebagian besar ukuran kinerja digunakan dalam masalah klasifikasi berdasarkan confusion matrix.

Ukuran pengklasifikasi yang paling sering dipaparkan adalah akurasi. Ukuran ini mengevaluasi keseluruhan efisiensi sebuah algoritma. Namun, akurasi dapat menyesatkan ukuran evaluasi ketika data tidak seimbang. Hal ini terjadi karena lebih banyak bobot ditempatkan di kelas mayoritas daripada di kelas minoritas sehingga lebih sulit bagi pengklasifikasi untuk tampil baik pada kelas minoritas. Sensitivity mengukur akurasi positif kasus sedangkan specificity mengukur keakuratan kasus

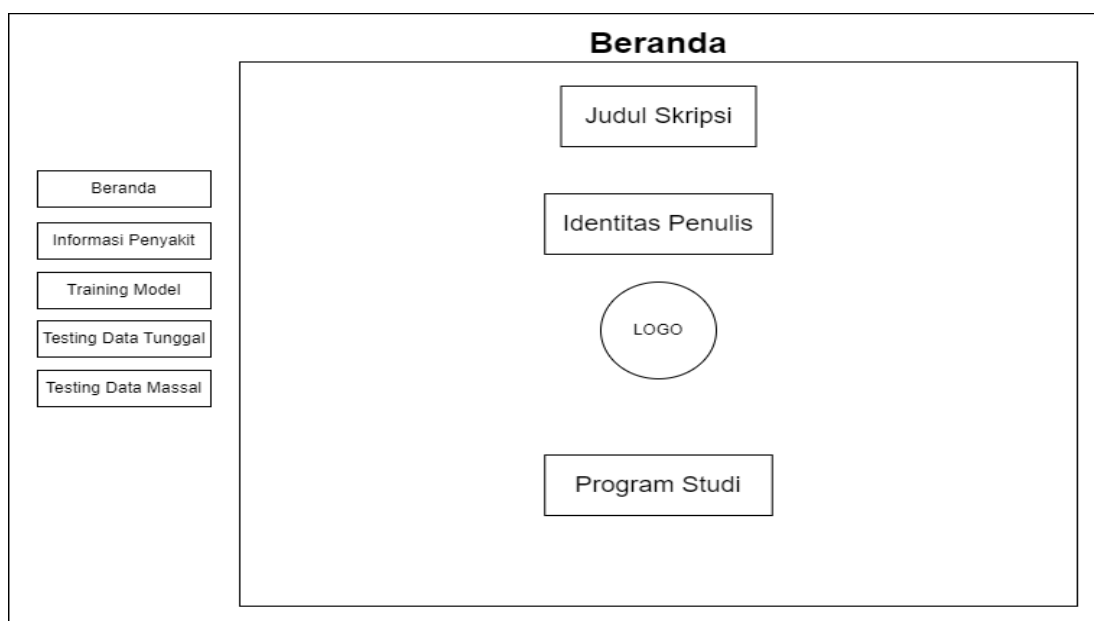
negatif. Sensitivity/Recall menilai efektivitas pengklasifikasi pada kelas positif/minoritas sementara spesificity menilai pengklasifikasi efektivitas pada kelas negatif/mayoritas. Untuk setiap analisis yang diberikan, biasanya ada trade-off antara sensitivity dan spesificity-nya. Precision adalah ukuran ketepatan model. Nilai precision yang lebih tinggi untuk classifier merupakan indikasi classifier yang baik.

3.5. Perancangan Sistem

Pada perancangan tahap ini, akan dijelaskan tentang perancangan sistem klasifikasi *Rontgen* dada menggunakan metode HOG dan *Random forest* berbasis *desktop*, dimana pada tahap ini akan ditampilkan perancangan menu dan antarmuka dengan tujuan agar aplikasi dapat dijalankan dengan mudah oleh pengguna, dan pada akhir perancangan sistem akan memberikan respon tentang jenis radiologi paru pada setiap gambar.

3.5.1. Rancangan Antarmuka Halaman Awal

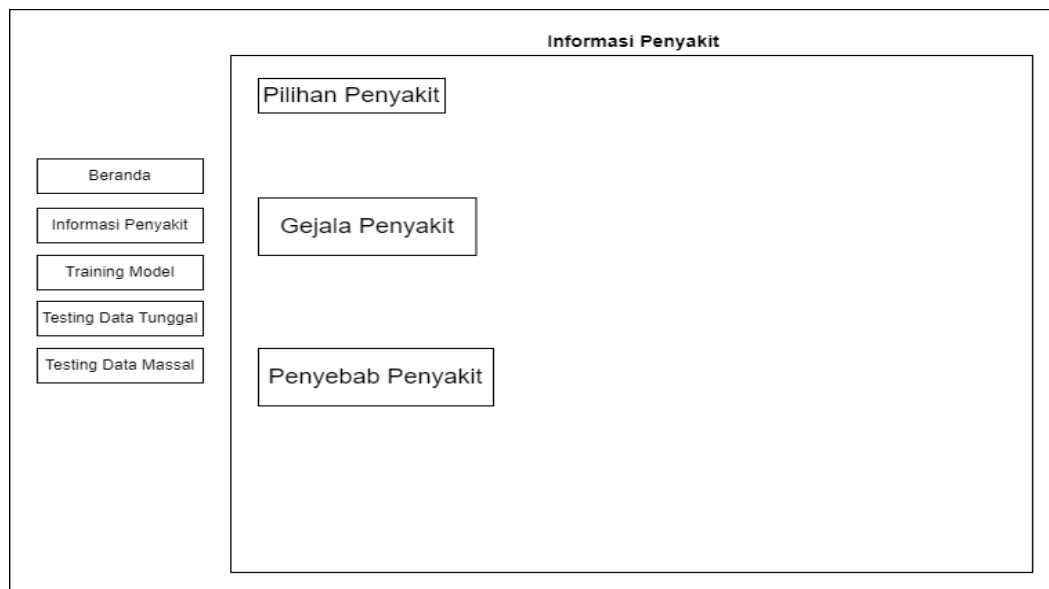
Pada halaman awal tampilan aplikasi yang pertama kali dilihat adalah tampilan judul penelitian, logo instansi, identitas penulis dan tombol untuk data *training* dan *testing*. Ketika tombol data *training* dan *testing* dipilih maka akan menampilkan halaman untuk memasukan berkas gambar sesuai dengan perintah yang diperlukan pada halaman yang ditampilkan, dapat dilihat pada gambar 3.2.



Gambar3.2 Tampilan Awal Desain

3.5.2. Desain Tampilan Informasi Penyakit

Di bagian sidebar kiri, terdapat menu sebagai navigasi untuk memudahkan pengguna berpindah antar halaman, yaitu halaman beranda, informasi penyakit, *training* model, *testing* data tunggal dan *testing* data massal. Pada halaman informasi penyakit dapat menampilkan gejala dan penyebab penyakit apabila user memilih nama penyakit yang ingin dilihat pada gambar 3.3.



Gambar3.3 Tampilan Informasi Penyakit

3.5.3. Desain *Training* Model

Pada halaman *training* model terdapat pemilihan data untuk melatih seluruh data latih, kemudian terdapat tombol pilih folder untuk memasukkan folder gambar yang ingin dilatih kemudian diharuskan untuk mengatur parameter nilai $n_estimators$ dan Max *features* nya, setelah itu terdapat tombol latih data untuk mengetahui status pada proses data yang telah dilatih dapat dilihat pada gambar 3.4.

Gambar3.4 Tampilan Training Model

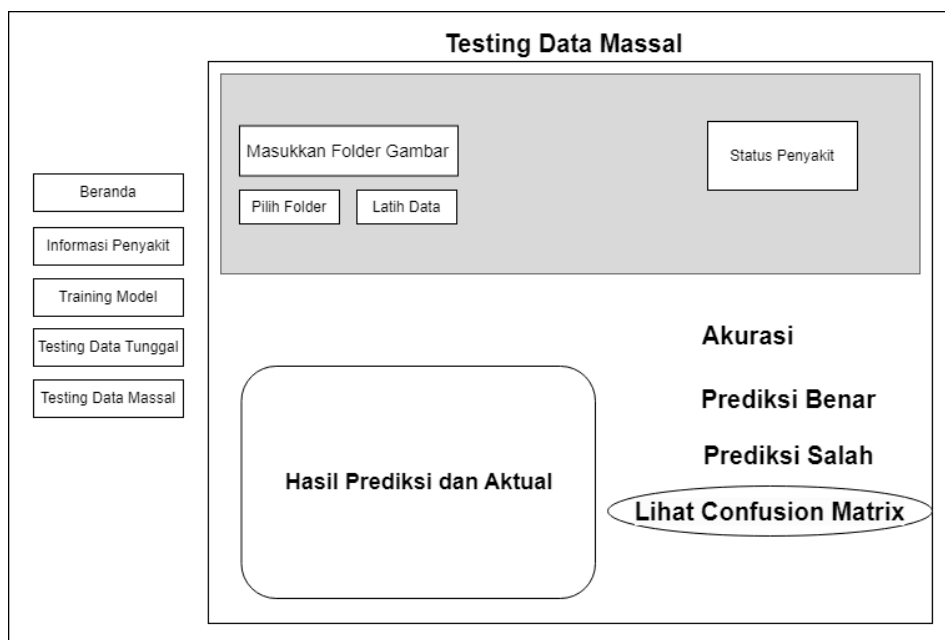
3.5.4. Desain *Testing Data Tunggal*

Pada halaman *testing* data tunggal menampilkan tombol untuk memasukkan folder gambar data tunggal yang ingin diinput, kemudian terdapat tombol proses data untuk menampilkan proses gambar seperti *citra asli*, *grayscale*, *sharpening*, *filtering*, *equalization_histogram*, dan *HOG*, lalu terdapat tombol latih data untuk mengetahui hasil dari penyakit yang telah diinput dapat dilihat pada gambar 3.5.

Gambar3.5 Tampilan Testing Data Tunggal

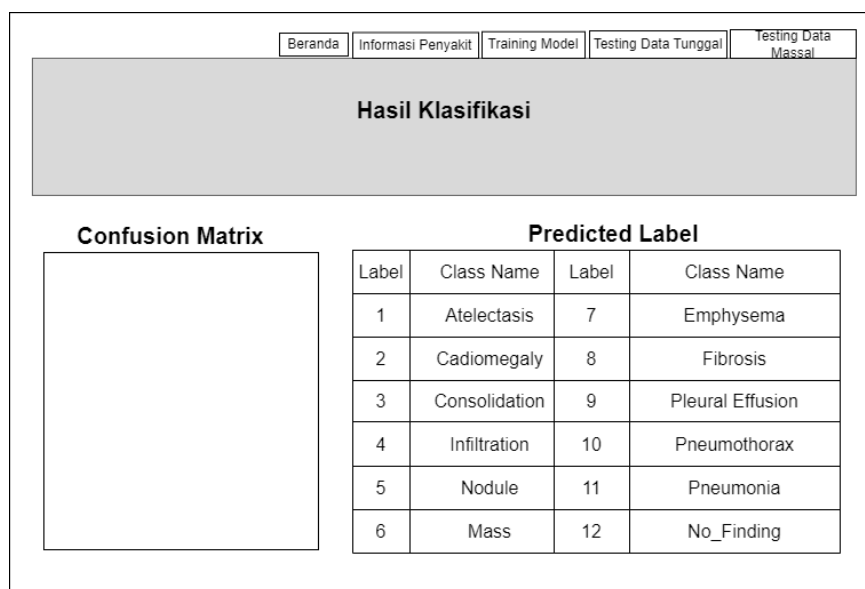
3.5.5. Desain *Testing* Data Massal

Pada halaman desain *testing* data massal kemudian jika ingin memasukkan folder gambar terdapat tombol pilih folder, lalu terdapat tombol latih data untuk proses menampilkan hasil akurasi, benar dan salah dari proses *testing* yang telah diinput, serta dengan akan muncul tabel benar dan salah sesuai hasil prediksi dapat dilihat pada gambar 3.6.



Gambar3.6 Tampilan Testing Data Massal

3.5.6. Desain Hasil Confusion Matrix



Gambar 3.7 Tampilan Hasil Klasifikasi Confusion Matrix

BAB 4

IMPLEMENTASI DAN PENGUJIAN

4.1. Implementasi Sistem

4.1.1 Spesifikasi kebutuhan Hardware

Dalam penelitian ini, menggunakan spesifikasi perangkat keras yaitu dapat dilihat pada table 4.1.

Tabel 4. 1 Spesifikasi kebutuhan Hardware

No	Jenis Komponen	Komponen yang digunakan
1	Processor	1. AMD A8 PRO-7150B R5,10 Compute Cores 4C+6G 1.90 GHz
2	System type	64-bit operating system, x64-based processor
3	Memory (RAM)	8,00 GB (6,95 GB usable)
4	Penyimpanan	500 GB SSD

4.1.2 Spesifikasi kebutuhan Software

Dalam penelitian ini, digunakan perangkat lunak dengan spesifikasi dapat dilihat pada table 4.2.

Tabel 4. 2 Spesifikasi kebutuhan Software

No	Jenis Komponen	Komponen yang digunakan
1	Python r	3.8.6-amd64
2	Sistem Operasi	1. Windows 10 64 bit 2. Windows Explorer 3. Microsoft Word 2010 4. Microsoft Edge 114.0.1823.37

3	<i>Library bahasa python</i>	a. <i>Imageio, Matplotlib,</i>
		b. <i>Matplotlib-inline,</i>
		c. <i>Numpy,</i>
		d. <i>Opencv-python,</i>
		e. <i>Pandas,</i>
		f. <i>Pillow,</i>
		g. <i>Scikit-image,</i>
		h. <i>Scikit-learn,</i>
		i. <i>Scipy,</i>
		j. <i>Seaborn dan</i>
		k. <i>Pandastable..</i>
		l. <i>Matplotlib</i>
4	<i>Visual Studio</i>	1. <i>Code -x64-1.70.2</i>

4.1.3 Implementasi Data

Dataset yang digunakan pada penelitian ini berupa citra digital toraks dan label. Dataset didapatkan melalui platform Kaggle.com, yang dikumpulkan oleh *vingroup Big Date Institute* (VinBigData). Dari data yang terkumpul terdapat 12 objek yang akan di klasifikasi yaitu *Atelectasis, Cardiomegaly, Concolidation, Effusion, Emphysema, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Fibrosis, No-Finding*. Jumlah citra digital yang dikumpulkan sebanyak 1.920 gambar sebagai data *training* dan 480 sebagai data *testing*, lalu untuk jumlah label dapat dilihat pada table 4.3.

Tabel 4. 3 Pembagian data training berdasarkan label

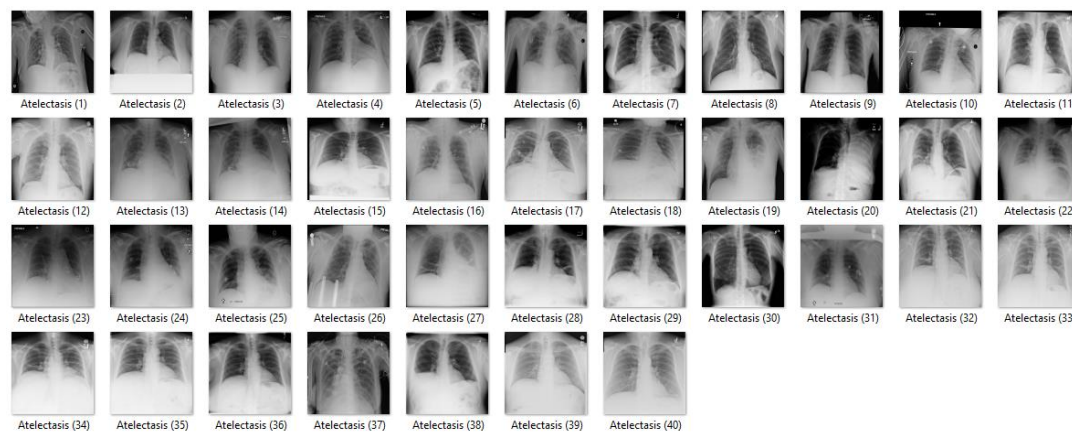
NO.	Jenis Objek	Jumlah Data
1.	<i>Atelectasis</i>	160
2.	<i>Cardiomegaly</i>	160
3.	<i>Concolidation</i>	160
4.	<i>Infiltration</i>	160
5.	<i>Nodule</i>	160

6.	<i>Mass</i>	160
7.	<i>Emphysema</i>	160
8.	<i>Fibrosis</i>	160
9.	<i>Pleural effusion</i>	160
10.	<i>Pneumothorax</i>	160
11.	<i>Pneumonia</i>	160
12.	<i>No_Finding</i>	160
Jumlah Data		1.920

Tabel 4. 4 Pembagian data testing berdasarkan label

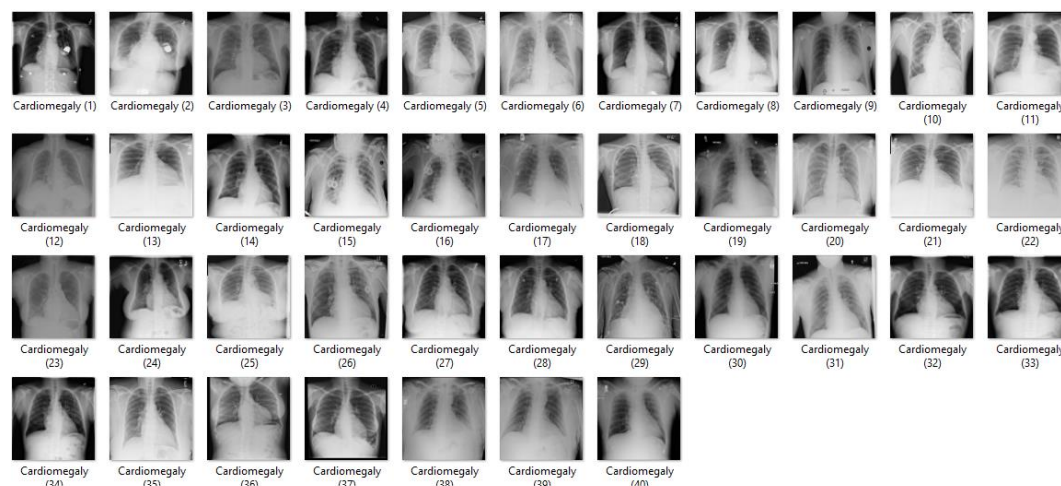
NO.	Jenis Objek	Jumlah Data
1.	<i>Atelectasis</i>	40
2.	<i>Cardiomegaly</i>	40
3.	<i>Concolidation</i>	40
4.	<i>Infiltration</i>	40
5.	<i>Nodule</i>	40
6.	<i>Mass</i>	40
7.	<i>Emphysema</i>	40
8.	<i>Fibrosis</i>	40
9.	<i>Pleural effusion</i>	40
10.	<i>Pneumothorax</i>	40
11.	<i>Pneumonia</i>	40
12.	<i>No_Finding</i>	40
Jumlah Data		480

Contoh citra *Rontgen* dada, berikut merupakan contoh data penyakit paru dapat dilihat pada Gambar 4.1. yang merupakan penyakit Atelectasis, yaitu penyakit paru dimana *alveolus* (struktur anatomi yang mempunyai bentuk berongga) yang tidak terisi oleh udara.



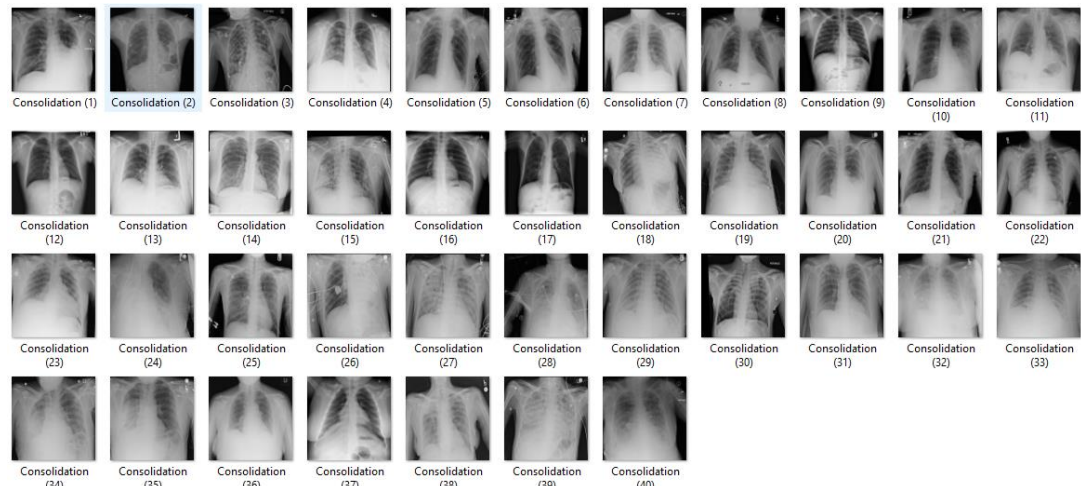
Gambar4.1 Data penyakit Atelectasis

Pada gambar berikut merupakan penyakit cardiomegaly yang merupakan merupakan kondisi dimana pembesaran yang tidak biasa terhadap kondisi jantung yang tidak memiliki ukuran jantung pada umumnya.



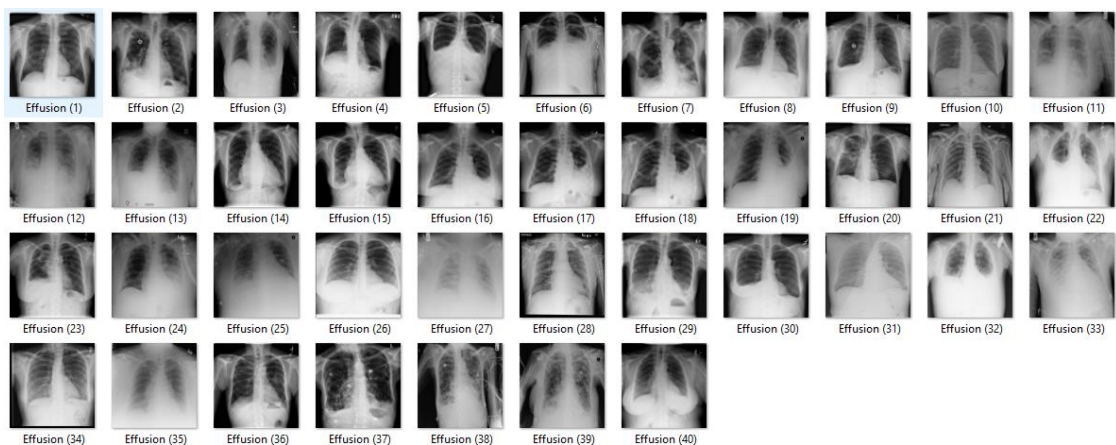
Gambar4.2 Data penyakit Cardiomegaly

Pada gambar berikut merupakan penyakit consolidation yaitu kondisi yang terjadi apabila terdapat sebuah cairan yang dapat mengisi saluran udara kecil pada paru-paru.



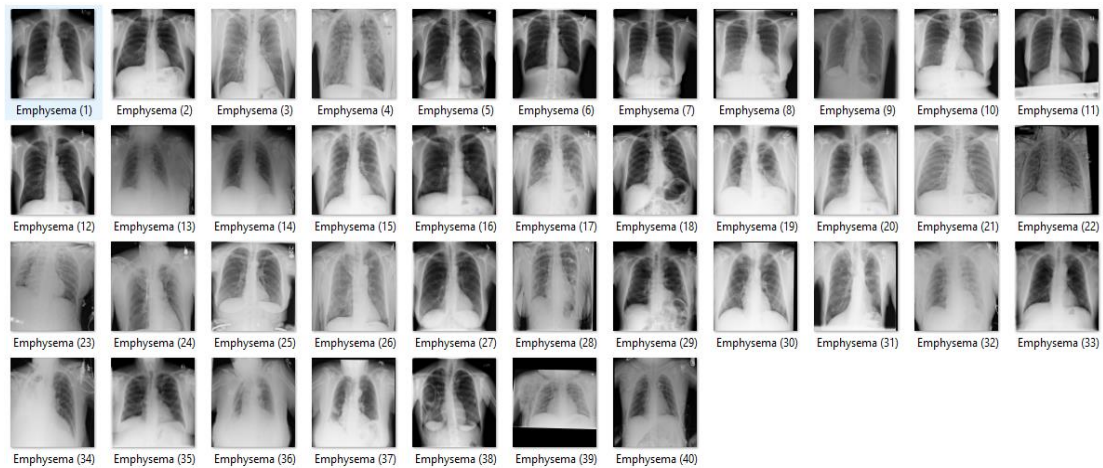
Gambar4.3 Data penyakit Consolidation

Gambar 4.4 berikut merupakan jenis penyakit Effusion yang diakibatkan oleh penumpukan cairan pada rongga pleuranya, pada rongga pleura terdapat lapisan yang menempel pada dinding yang berada pada rongga dada dan membungkus paru-paru.



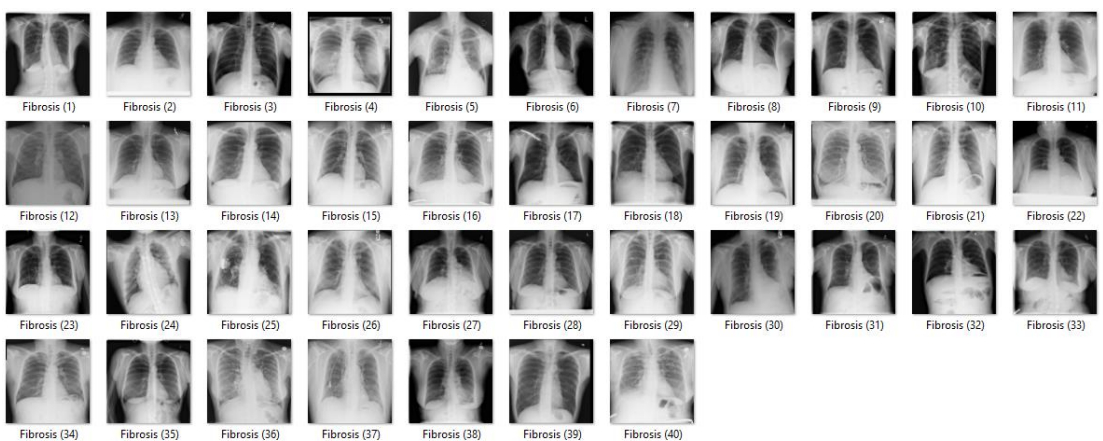
Gambar4. 4 Data penyakit Effusion

Gambar 4.5 merupakan Penyakit ini disebabkan oleh kerusakan *alveoli*, kantung udara pada paru-paru, memiliki fungsi sebagai penampung untuk pertukaran udara oksigen dengan karbondioksida.



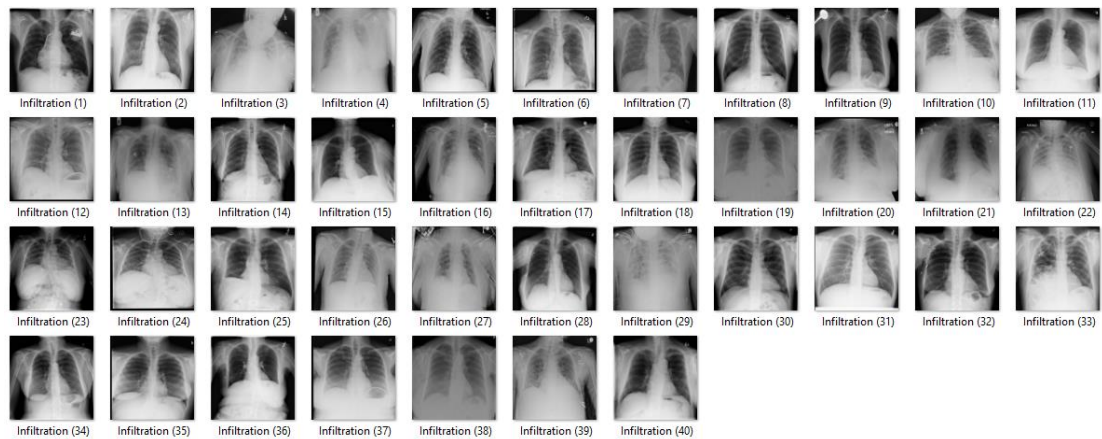
Gambar4. 5 Data penyakit Emphysema

Pada gambar 4.6 merupakan penyakit fibrosis yang terjadi ketika jaringan paru-paru rusak atau terluka. Jaringan yang kaku dan menebal dapat membuat lebih sulit untuk paru-paru bekerja.



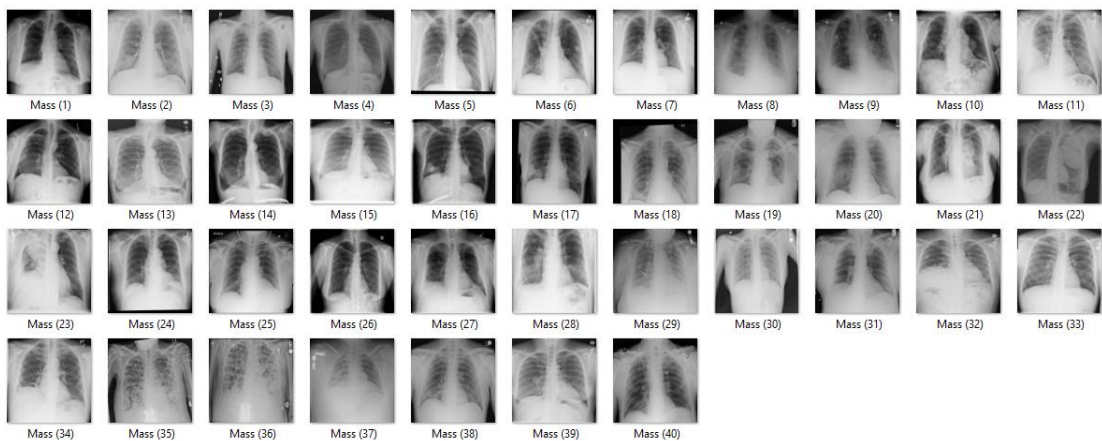
Gambar4. 6 Data penyakit Fibrosis

Pada gambar 4.7 merupakan penyakit *Infiltration* pada paru-paru biasanya sering berhubungan dengan penyakit *Pneumonia* dan *tuberculosis*. *Infiltration* adalah penyakit seperti nanah, darah, atau protein, seperti zat yang lebih padat daripada udara yang berada didalam paru.



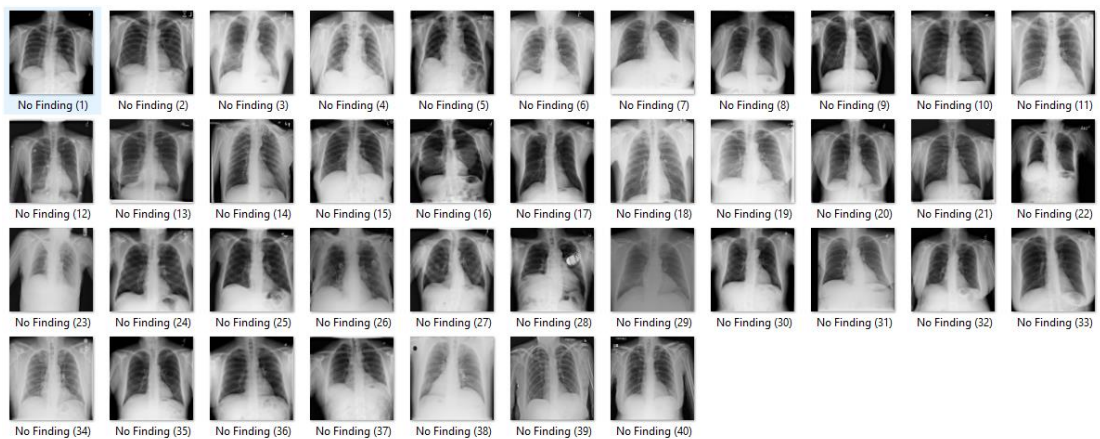
Gambar4.7 Data penyakit Infiltration

Gambar 4.8 berikut merupakan Penyakit mass hampir sama dengan nodule hanya saja ukurannya yang membedakannya, dan penyakit *mass* memungkinkan memiliki kelainan yang lebih ganas dari pada nodule .



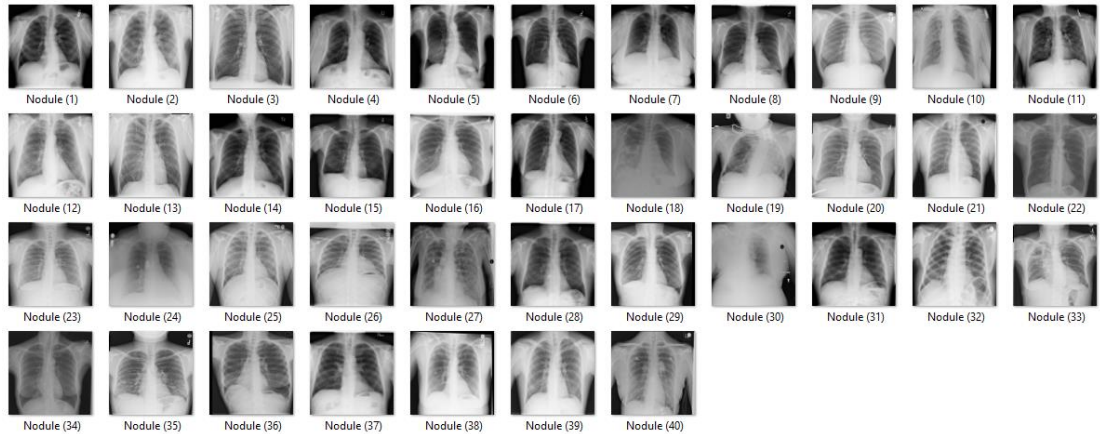
Gambar4.8 Data penyakit Mass

Gambar 4.9 merupakan No_Finding yaitu tidak ada gejala penyakit apapun, sehingga yang dihasilkan bukanlah sebuah penyakit.



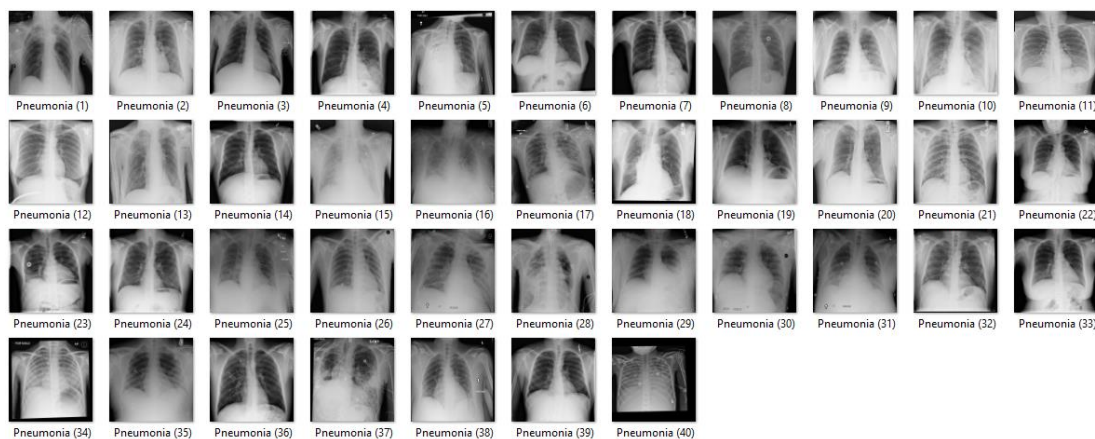
Gambar4.9 Data No_finding

Pada gambar 4.10 merupakan Nodule yaitu terdapat benjolan yang terbentuk di dalam paru-paru. Biasanya, *nodule* paru-paru ditemukan secara tidak sengaja melalui pemeriksaan radiologi seperti *Rontgen* dada atau tomografi komputer (CT scan).



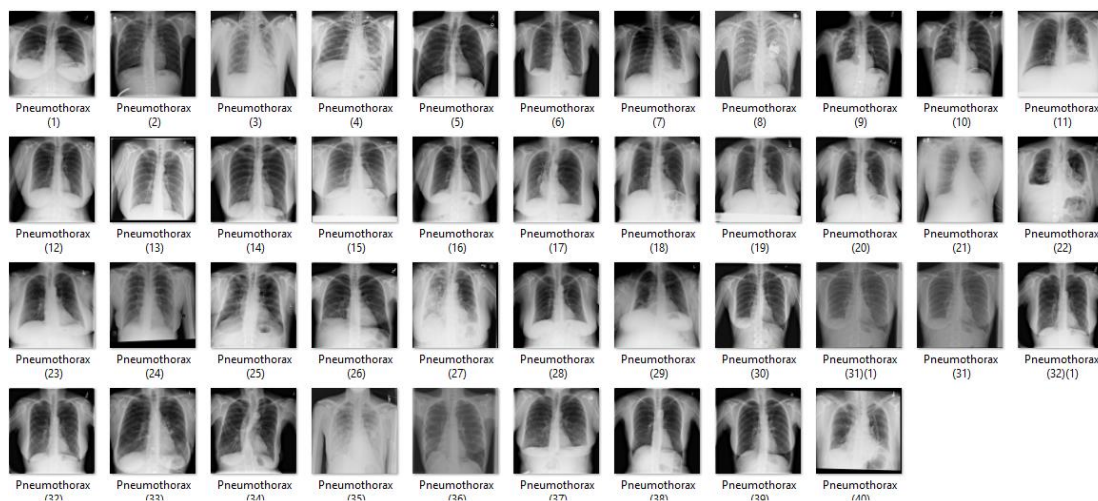
Gambar4.10 Data penyakit Nodule

Pada gambar 4.11 merupakan penyakit Pneumonia yaitu suatu kondisi peradangan pada paru-paru yang disebabkan oleh infeksi. Gejala *pneumonia* bisa bervariasi, mulai dari ringan hingga berat.



Gambar4.11 Data penyakit Pneumonia

Pada gambar 4.12 terdapat jenis penyakit Pneumothorax yang merupakan jenis penyakit yang diakibatkan diakibatkan karena adanya cedera yang terjadi pada dinding atau robekan jaringan paru-paru. Dampaknya dapat terjadi dan menyebabkan paru-paru mengempis dan tidak bias mengembang.



Gambar4.12 Data penyakit Pneumothorax

4.1.4. Implementasi HOG

Dilakukan proses ekstraksi ciri menggunakan *Histogram of Oriented Gradient* untuk fitur gambar. *Histogram Gradien Oriented*, atau HOG, adalah metode yang biasanya digunakan dalam aplikasi pengenalan objek untuk mengekstraksi fitur dari gambar. Metode ini melibatkan beberapa langkah, seperti pra-pemrosesan gambar, perhitungan gradien, pembagian sel, normalisasi blok, dan pembuatan histogram. Dalam pemrosesan citra RGB, citra dapat diubah menjadi citra grayscale sebagai langkah awal. Selanjutnya, citra *grayscale* ini dapat digunakan untuk melakukan ekstraksi ciri menggunakan metode HOG. Setelah proses ekstraksi ciri selesai, hasil yang diperoleh dari tahap pengujian (*testing*) akan dibandingkan dengan hasil model yang diperoleh dari klasifikasi pada tahap pelatihan (Ningrum, et al 2021).

4.1.5. Implementasi Random forest

Setelah melewati proses klasifikasi dengan menggunakan metode *Random forest*, langkah ini melibatkan penggunaan metode ekstraksi HOG atau vektor fitur yang diperoleh dari pelatihan dengan menggunakan ekstraksi fitur HOG. Hasil dari tahap ini akan menghasilkan klasifikasi dalam bentuk objek abnormal dari hasil radiologi paru-paru. Objek abnormal akan memilih kelas dengan tingkat kepopuleran tertinggi. Berdasarkan output ini, hasil akhir dari klasifikasi akan berupa objek nama dari hasil radiologi *thorax*.

Random forest adalah teknologi yang merepresentasikan sekumpulan metode pembelajaran ensemble untuk klasifikasi random yang berfungsi dengan mengambil keputusan menggunakan banyak suara pohon dan memprediksi fitur data sebagai berikut: hasil klasifikasi dari setiap pohon dikumpulkan untuk gambar input, setelah itu, suara terbanyak dikumpulkan untuk memberikan label kelas yang dihasilkan. *Random forest* (RF) adalah teknik kecerdasan buatan dan metode modern yang kuat untuk klasifikasi data dan pemodelan. RF telah diterapkan untuk membandingkan tempat yang diekstraksi (fitur vektor) dari tahap pelatihan dan pengujian agar sesuai dengan yang bersangkutan (Huda Mady, et al. 2018).

Parameter Algoritma *Random forest* dapat diubah untuk meningkatkan kinerja model yang dibangun. Untuk meningkatkan kinerja model maka *hyperparameter* ini dapat diubah. *Hyperparameter* algoritma *Random forest* yang digunakan dalam

penelitian ini adalah: Setiap kombinasi parameter memiliki *hyperparameter* yang digunakan dalam penelitian ini adalah sebagai berikut:

1. *N-estimators*: Pada penelitian ini, penulis menggunakan parameter *n_estimators* dari 10 hingga 1000. Jumlah pohon yang akan dibangun dalam *Random forest* ditetapkan oleh parameter ini. Semakin banyak pohon yang dibangun, semakin baik kemampuan model untuk mempelajari pola yang kompleks, tetapi akan lebih lama untuk melatih model. Kebijakan umum adalah memilih jumlah pohon yang cukup besar untuk mencapai konvergensi performa model.
2. *Max-Features*: Parameter ini menentukan jumlah fitur yang dipertimbangkan saat mencari pemisahan terbaik di setiap simpul. Pilihan umum termasuk "sqrt" (menggunakan akar kuadrat dari jumlah fitur), "log2" (menggunakan log basis 2 dari jumlah fitur), atau bilangan bulat (menggunakan jumlah fitur tertentu).

4.1.6. Implementasi Perancangan Aplikasi

Implementasi perancangan aplikasi antarmuka berdasarkan rancangan system yang sudah dijelaskan pada Bab 3 sebelumnya yaitu sebagai berikut:

a. Tampilan Pada Halaman Awal Aplikasi

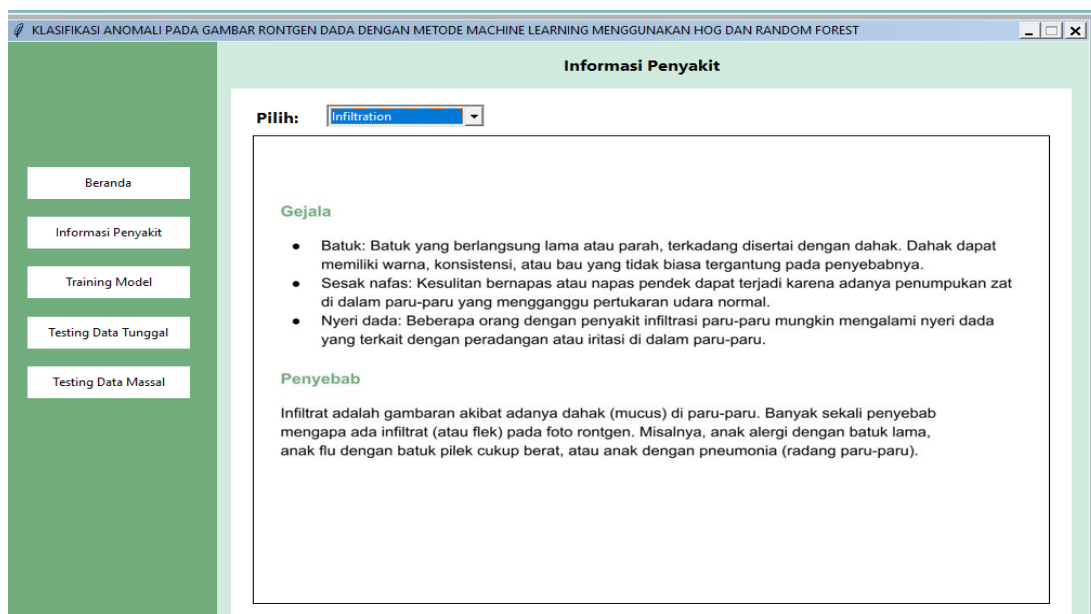
Tampilan awal aplikasi adalah tampilan yang muncul saat pertama kali aplikasi dibuka, pada aplikasi ini terdapat judul tugas akhir, identitas pembuat aplikasi dan tombol data *testing* serta data *training*. Pada tombol data *training* dan data *testing* adalah langkah awal untuk memulai proses klasifikasi citra *Rontgen* dada. Tampilan halaman depan dapat dilihat pada Gambar 4.13.



Gambar4.13 Implementasi Tampilan Awal

b. Tampilan Aplikasi Informasi Penyakit

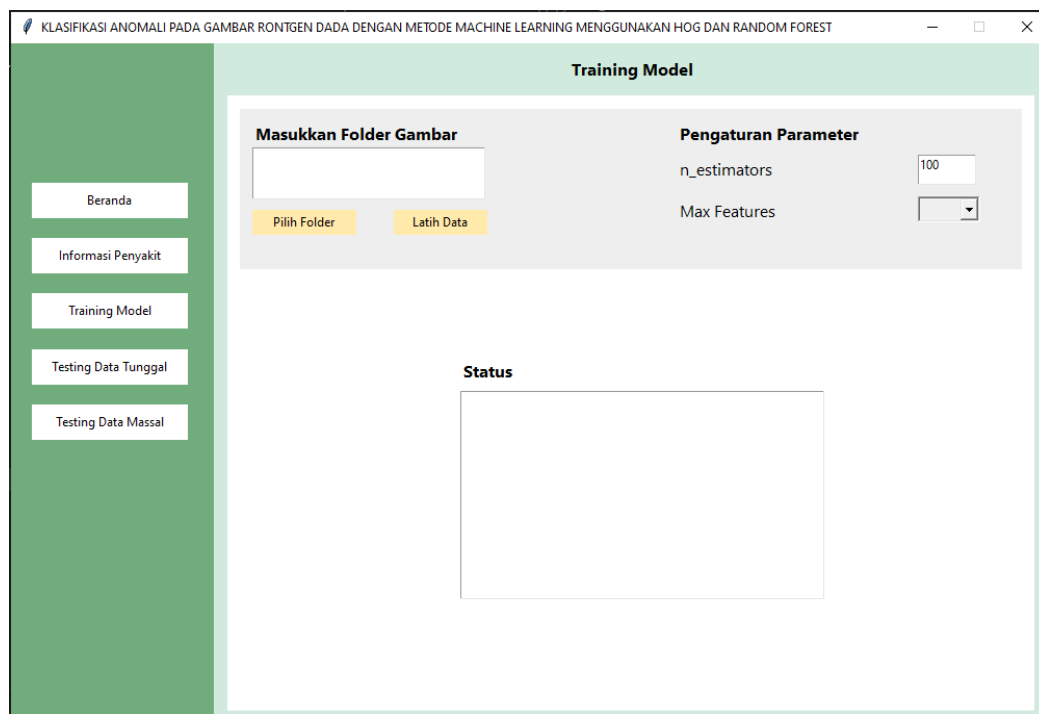
Merupakan tampilan informasi 12 penyakit paru dengan keterangan gejala serta penyebab dari penyakit tersebut dapat dilihat pada Gambar 4.14.



Gambar4.14 Implementasi Tampilan Informasi Penyakit

c. Tampilan Aplikasi Halaman data *Training Model*

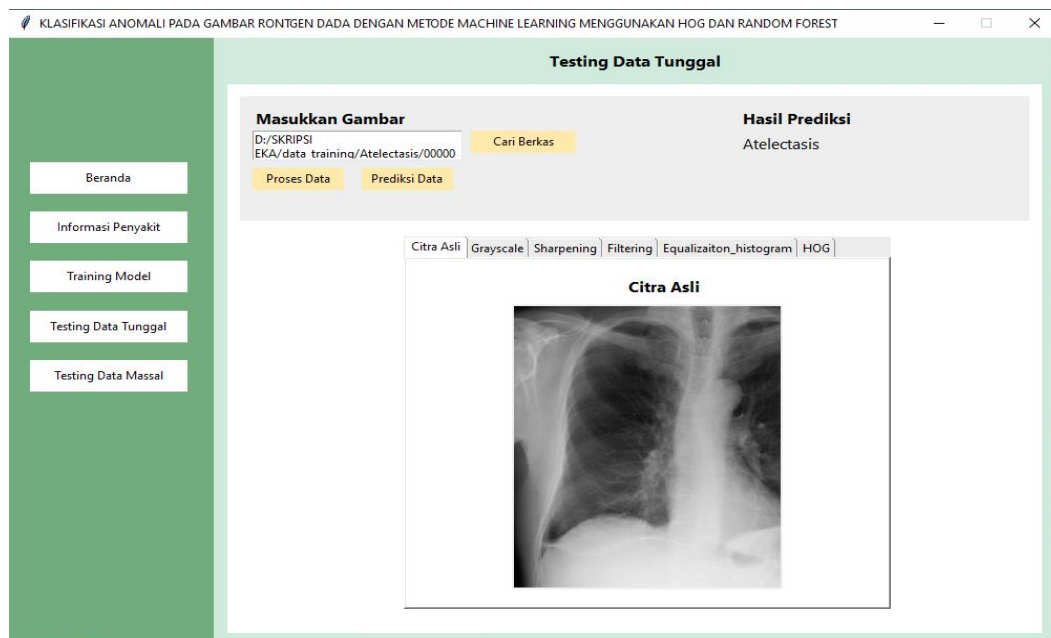
Tampilan aplikasi pada halaman *training* terdiri dari tombol beranda, tombol data latih, tombol memilih folder, dan tombol data *testing*. Setelah data gambar dimasukkan maka akan muncul hasil pada kolom Status dan dapat di atur pada pengaturan parameternya yaitu *n_estimators* dan *Max features*, lalu pada bagian bawah terdapat kolom untuk melihat hasil dari data *training*. Perancangan ini dapat dilihat pada Gambar 4.15.



Gambar4.15 Implementasi Tampilan Data Training

d. Tampilan Aplikasi Halaman *Testing Data Tunggal*

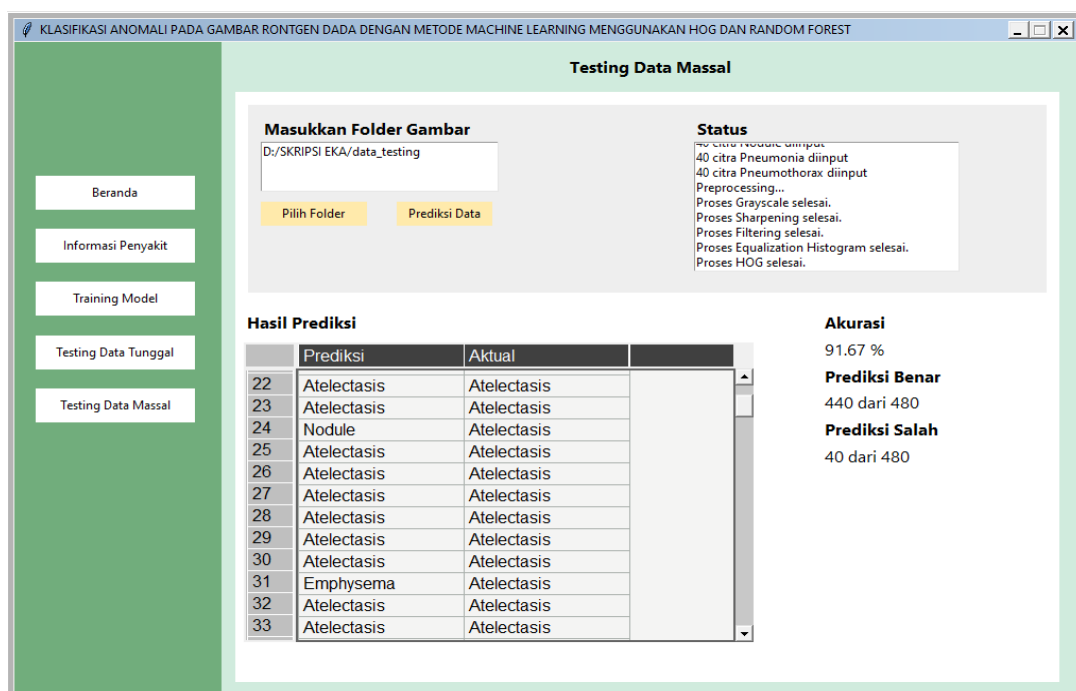
Tampilan aplikasi *testing* digunakan untuk melakukan pengujian pada data klasifikasi *Rontgen* dada dengan menggunakan metode *Histogram of Oriented Gradient*. Pada bagian ini terdapat tombol untuk mencari berkas/data sehingga data dapat di input kedalam aplikasi. Kemudian saat menekan tombol proses data maka data akan muncul pada kolom gambar, sehingga dapat melihat data dari pilihan citra asli, *grayscale*, *sharpening*, *filtering*, *equalization_histogram*, dan HOG. Kemudian pada tombol prediksi data maka akan muncul pada kolom hasil prediksi dan kecepatan perhitungan datanya pada kolom Run Time. Tampilan halaman *testing* ini dapat dilihat pada Gambar 4.16.



Gambar4.16 Implementasi Tampilan Data Testing

e. Tampilan Aplikasi Halaman *Testing* Data Massal

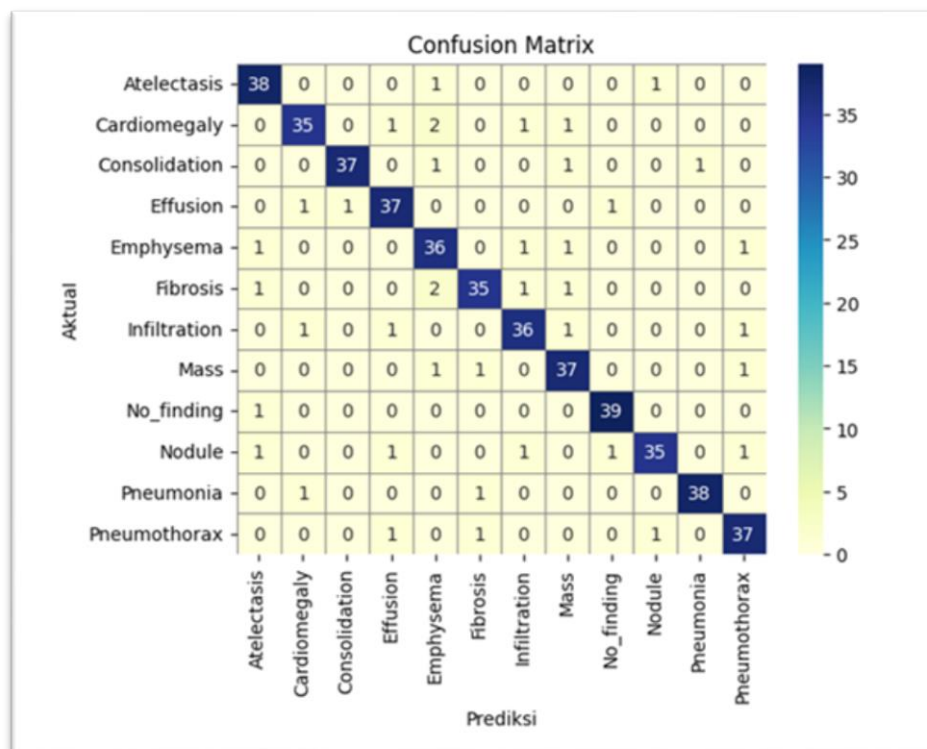
Selanjutnya halaman yang menampilkan hasil dari proses klasifikasi setelah gambar diinput. Pada halaman ini telah ditampilkan status gambar yang telah di *testing*, hasil prediksi, jumlah prediksi benar, jumlah prediksi salah, dan tabel daftar objek abnormal yang terklasifikasi setiap objeknya. Tampilan hasil dapat dilihat pada Gambar 4.17.



Gambar4.17 Implementasi Tampilan data Testing

f. Tampilan Hasil Akurasi dengan Confusion Matrix

Hasil dari model yang telah diproses pada *testing* data massal tersebut selanjutnya akan dievaluasi dengan menggunakan *Confusion Matrix*. Hasil yang akan ditampilkan yaitu berupa tabel *confusion matrix* sehingga memudahkan pengguna untuk membaca hasil evaluasi saat mengklasifikasi gambar *rontgen*. Hasil evaluasi dapat dilihat pada Gambar 4.18d



Gambar 4.18 Evaluasi Confusion Matrix

4.2. Implementasi Model

Tabel 4. 5 Implementasi Model dengan Data Testing

n_estimator (Jumlah pohon)	<i>Max features</i> (Kedalaman pohon)	Jumlah Data Benar	Jumlah Data Salah	Akurasi
10	Sqrt	349	131	72.71 %
	log2	313	167	65.21 %
20	Sqrt	349	131	72.71 %
	log2	312	168	65.00 %
30	Sqrt	366	114	76.25 %
	log2	316	164	65.83 %
40	Sqrt	359	121	74.79 %
	log2	321	159	66.88 %
50	Sqrt	370	110	77.08 %
	log2	322	158	67.08 %
60	Sqrt	381	99	79.38 %
	log2	322	158	67.08 %
70	Sqrt	390	90	81.25 %
	log2	328	152	68.33 %
80	Sqrt	398	82	82.92 %
	log2	323	157	67.29 %
90	Sqrt	409	71	85.21 %
	log2	323	157	67.29 %
99	Sqrt	434	46	90.42 %
	log2	320	160	66.67 %

100	Sqrt	440	40	91.67 %
	log2	322	158	67.08 %
101	Sqrt	434	46	90.42 %
	log2	322	158	67.08 %
110	Sqrt	416	64	86.67 %
	log2	320	160	66.67 %
300	Sqrt	366	114	76.25 %
	log2	323	157	67.29 %
1000	Sqrt	356	124	74.17 %
	log2	347	133	72.29 %

4.3. Evaluasi

4.3.1. Hasil Evaluasi Model

Evaluasi dilakukan pada seluruh model kombinasi yang telah melewati *tuning* parameter. Metode evaluasi berfungsi sebagai patokan untuk menghitung kinerja model dalam mengklasifikasi *rontgen* paru dengan menggunakan *Random Forest* dan HOG (*Histogram Of Oriented Gradient*). Penelitian ini menggunakan *confusion matrix* untuk mengukur nilai akurasi 12 penyakit paru dan dua parameter tambahan yaitu *Random Forest* dan HOG.

Informasi hasil dari semua matriks penilaian ditampilkan dengan menggunakan visualisasi *heatmap* dengan pola secara berurut yaitu TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*). Visualisasi *heatmap* dapat dilihat pada Gambar 4.18

Tabel 4. 6 Nilai TP, TN, FP dan FN pada Data Testing

	TP	TN	FP	FN	Total
<i>Atelectasis</i>	38	35	4	2	92
<i>Cardiomegaly</i>	35	37	3	5	90
<i>Concolidation</i>	37	37	1	3	95
<i>effusion</i>	37	36	4	3	90
<i>Emphysema</i>	36	35	7	4	87
<i>Fibrosis</i>	35	36	3	5	90
<i>Infiltration</i>	36	37	4	4	90
<i>Mass</i>	37	39	5	3	90
<i>No_Finding</i>	39	35	2	1	96
<i>Nodule</i>	35	38	2	5	91
<i>Pneumonia</i>	38	37	1	2	96
<i>Pneumothorax</i>	37	38	4	3	91

Berdasarkan hasil klasifikasi dari data *testing* yang tertuang dalam Tabel 4.7, maka evaluasi dapat dilakukan dengan menghitung nilai pada *rontgen* paru dan dua parameter tambahan yaitu *Random Forest* dan HOG dengan menggunakan persamaan yang telah dijelaskan pada bab sebelumnya:

$$\text{Atelectasis} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{38+35}{38+4+35+2} \times 100\% = 92\%$$

$$\text{Cardiomegaly} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{35+37}{35+3+37+5} \times 100\% = 90\%$$

$$\text{Concolidation} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{37+37}{37+1+37+3} \times 100\% = 95\%$$

$$\text{effusion} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{37+36}{37+4+36+3} \times 100\% = 91\%$$

$$\text{Emphysema} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{36+35}{36+7+35+4} \times 100\% = 87\%$$

$$\text{Fibrosis} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{35+36}{35+3+36+5} \times 100\% = 90\%$$

$$\text{Infiltration} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{36+37}{36+4+37+4} \times 100\% = 90\%$$

$$\begin{aligned}
\textit{Mass} &= \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{37+39}{37+5+39+3} \times 100\% = 90\% \\
\textit{No_Finding} &= \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{39+35}{39+2+35+1} \times 100\% = 96\% \\
\textit{Nodule} &= \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{35+38}{35+2+38+5} \times 100\% = 91\% \\
\textit{Pneumonia} &= \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{38+37}{38+1+37+2} \times 100\% = 96\% \\
\textit{Pneumothorax} &= \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{37+38}{37+4+38+3} \times 100\% = 91\% \\
\textit{Akurasi} &= \frac{92+90+95+91+87+90+90+90+96+91+96+91}{12} \times 100\% = 91,67\%
\end{aligned}$$

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari hasil pengujian terhadap sistem klasifikasi *Rontgen* dada menggunakan *Histogram of Oriented Gradient* dapat diperoleh kesimpulan yaitu sebagai berikut:

1. *Training* dilakukan sebanyak 30 kali dengan variasi nilai parameter *n_estimators* dan *Max features* yang berbeda-beda sehingga menghasilkan nilai terbaiknya.
2. Data yang dihasilkan dalam penelitian ini termasuk kedalam kategori baik karena menghasilkan akurasi tertinggi yaitu menggunakan algoritma klasifikasi *Random Forest* dengan hasil akurasi sebesar 92%.
3. Sistem berbasis desktop yang dibangun dalam penelitian ini memudahkan pengguna untuk memeriksa kondisi abnormal melalui gambar x-ray toraks dengan mudah dan dengan proses yang cepat.

5.2. Saran

Dari hasil penelitian ini, memiliki beberapa saran yang dapat dipertimbangkan untuk mengembangkan sistem dan penelitian kedepannya:

1. Menerapkan transformasi pengolahan citra yang lain dan menambah variasi transformasi pada data *augmentation* meningkatkan variasi dataset dan meningkatkan akurasi.
2. Diperlukan algoritma *deep learning* lainnya untuk mengklasifikasi objek.
3. Dapat memberikan penilaian secara kasat mata pada citra, agar gambar dapat dibedakan pada setiap penyakit.
4. Dapat menerapkan konsep ansambel model dengan mengkombinasikan model dari beberapa algoritma untuk meningkatkan hasil prediksi.

DAFTAR PUSTAKA

- Iskandar, D.A . 2021. Klasifikasi dan Klasifikasi Keretakan Jalan Menggunakan Metode *You Only Look Once*. Skripsi. Universitas Sumatera Utara.
- Siregar, G.L. 2019. Identifikasi Penyakit Defisiensi Besi Berdasarkan Kelainan Darah Merah dengan Menggunakan metode *convolutional Neural Network*. Skripsi. Universitas Sumatera Utara.
- Bambang, P.H. (2021). Penerapan Convolutional Neural Network pada Citra *Rontgen dada* untuk Klasifikasi SARS-CoV-2. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 747–759. (Online) <https://doi.org/10.29207/resti.v5i4.3153>(diakses pada 18 Oktober 2021)
- Hreelaita, D.S. (2021). *Dokter: Kanker paru jadi penyebab kematian tertinggi di Indonesia*. Antaranews.Com. (Online) <https://www.antaranews.com/berita/2294506/dokter-kanker-paru-jadi-penyebab-kematian-tertinggi-di-indonesia> (diakses 4 November 2021)
- Jenggawah, N., S., Berpikir, K., Dan, K., & Belajar, M. 2010. *Digital Digital Repository Repository Universitas Universitas Jember Jember Digital Jember Digital Repository Repository Universitas Universitas Jember*. 68–74. (Online)
- Li, G., Song, Z., & Fu, Q. 2018. A New Method of Image Detection for Small Datasets under the Framework of YOLO Network. *Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018, October 2018*, 1031–1035. (Online) <https://doi.org/10.1109/IAEAC.2018.8577214> (diakses 22 Oktober 2021)
- Rahmadewi, R. 2018. Analisa Perbandingan Beberapa Metode Klasifikasi Tepi Pada Citra *Rontgen* Penyakit Paru - Paru. *Media Elektro*, VI(1), 9–12.(Online) <https://jurnal.um-palembang.ac.id/integrasi/article/viewFile/998/872> (diakses 18 Oktober 2021)

- Rahmadewi, R., Kurnia, R., Listyalina, L., Fauzi, A., Riana, D., Ugm, F., Prihasty, W., Kusumawardhani, A., & Siswanto. (2013). Metode Segmentasi Canny pada Citra *Rontgen* untuk Klasifikasi Penyakit Paru. *Jurnal Nasional Teknik Elektro*, 1(2), 140–145. (online)
<http://jti.respati.ac.id/index.php/jurnaljti/article/viewFile/1/1> (diakses 18 Oktober 2021)
- Shianto, K. A., Gunadi, K., & Setyati, E. 2019. Klasifikasi Jenis Mobil Menggunakan Metode YOLO Dan Faster R-CNN. *Jurnal Infra*, 7(1), 157–163. (Online)
<http://publication.petra.ac.id/index.php/teknik-informatika/article/view/8065> .
 (diakses 22 Oktober 2021)
- Sistem, P., Tracking, P., Hudaya, M. A., Santoso, I., & Adi, A. (2021). *PERHITUNGAN KENDARAAN PADA CITRA BERGERAK MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK*. 9(1). (Online)
<https://ejournal3.undip.ac.id/index.php/transient/article/view/26014>. (diakses 22 Oktober 2021)
- Utomo, P., , W., & Suryani, E. 2016. Sistem Diagnosa Penyakit Paru Berdasarkan Foto *Rontgen* Dengan Pendekatan Fuzzy Learning Vector Quantization. *Jurnal Teknologi & Informasi ITSmart*, 1(2), 102. (online).
<https://doi.org/10.20961/its.v1i2.604> (diakses 19 Oktober 2021)
- Suyatno, F. E. R. R. Y., Istofa, I., & Yuniarsari, L. (2007). Rekayasa Sistem Pengatur Parameter Pesawat Sinar-X Diagnostik Berbasis Mikrokontroller Keluarga MCS 51. *Semin Nas III, SDM Teknol Nukl. Yogyakarta: Sekolah Tinggi Teknologi Nuklir-BATAN*, 337-44.
- Dougherty, G. (2009). *Digital Image Processing for Medical Applications*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511609657
- Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11), 3212-3232

- S. Sedai, P. K. Roy and R. Garnavi, "*Right ventricle landmark detection using multiscale HOG and Random forest classifier*," *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, Brooklyn, NY, USA, 2015, pp. 814-818, doi: 10.1109/ISBI.2015.7163996.
- Ningrum. L. Siska, D. Muhammad, E.A. (2021) "Pengenal ASL Menggunakan Metode Ekstraksi HOG dan Klasifikasi *Random forest*".E-ISSN 2503-2933. doi:10.35957/jatisi.v8i2.456
- H. Mady and S. M. S. Hilles, "*Face recognition and detection using Random forest and combination of LBP and HOG features*," *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, Shah Alam, Malaysia, 2018, pp. 1-7, doi: 10.1109/ICSCEE.2018.8538377. (Diakses pada 15 Juni 2023)
- VinBigData Chest X-ray Abnormalities Detection* / Kaggle. (2020). Kaggle. <https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection/overview>



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI**
UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

Jalan Universitas No. 9A Gedung A, Kampus USU Medan 20155, Telepon: (061) 821007
Laman: <http://Fasilkomti.usu.ac.id>

KEPUTUSAN
DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI
NOMOR : 1561/UN5.2.14.D/SK/SPB/2024
DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA

- Membaca** : Surat Permohonan Mahasiswa Fasilkom-TI USU tanggal 17 Mei 2024 perihal permohonan ujian skripsi:
Nama : EKA WULANDARI
NIM : 171402084
Program Studi : Sarjana (S-1) Teknologi Informasi
Judul Skripsi : Klasifikasi Anomali Pada Gambar Rontgen Dada Dengan Metode Machine Learning Menggunakan Histogram of Oriented Gradient dan Random Forest
- Memperhatikan** : Bahwa Mahasiswa tersebut telah memenuhi kewajiban untuk ikut dalam pelaksanaan Meja Hijau Skripsi Mahasiswa pada Program Studi Sarjana (S-1) Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara TA 2023/2024.
- Menimbang** : Bahwa permohonan tersebut diatas dapat disetujui dan perlu ditetapkan dengan surat keputusan
- Mengingat** : 1. Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Peraturan Pemerintah Nomor 17 tahun 2010 tentang pengelolaan dan penyelenggara pendidikan.
3. Keputusan Rektor USU Nomor 03/UN5.1.R/SK/SPB/2021 tentang Peraturan Akademik Program Sarjana Universitas Sumatera Utara.
4. Surat Keputusan Rektor USU Nomor 1876/UN5.1.R/SK/SDM/2021 tentang pengangkatan Dekan Fasilkom-TI USU Periode 2021-2026
- MEMUTUSKAN**
- Menetapkan** :
- Pertama** : Membentuk dan mengangkat Tim Penguji Skripsi mahasiswa sebagai berikut:
- Ketua : Dedy Arisandi ST., M.Kom.
NIP: 197908312009121002
- Sekretaris : Seniman S.Kom., M.Kom.
NIP: 198705252014041001
- Anggota Penguji : Niskarto Zendrato S.Kom., M.Kom
NIP: 198909192018051001
- Anggota Penguji : Ulfi Andayani S.Kom., M.Kom
NIP: 198604192015042004
- Moderator : -
- Panitera : -
- Kedua** : Segala biaya yang diperlukan untuk pelaksanaan kegiatan ini dibebankan pada Dana Penerimaan Bukan Pajak (PNPB) Fasilkom-TI USU Tahun 2024.
- Ketiga** : Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan bahwa segala sesuatunya akan diperbaiki sebagaimana mestinya apabila dikemudian hari terdapat kekeliruan dalam surat keputusan ini.

Tembusan :

1. Ketua Program Studi Sarjana (S-1) Teknologi Informasi
2. Yang bersangkutan
3. Arsip

Medan, 20 Mei 2024

Ditandatangani secara elektronik oleh:
Dekan



Maya Silvi Lydia
NIP 197401272002122001