

**KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN CITRA
X-RAY MENGGUNAKAN METODE CONVOLUTIONAL NEURAL
NETWORK**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAMMAM AFIQ MURTADHO

18.11.1931

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN CITRA
X-RAY MENGGUNAKAN METODE CONVOLUTIONAL NEURAL
NETWORK**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAMMAM AFIQ MURTADHO

18.11.1931

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN CITRA X-
RAY MENGGUNAKAN METODE CONVOLUTIONAL NEURAL
NETWORK**

yang disusun dan diajukan oleh

Hammam Afiq Murtadho

18.11.1931

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Desember 2023

Dosen Pembimbing,



Windha Mega Pradnya D, M.Kom

NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN CITRA
X-RAY MENGGUNAKAN METODE CONVOLUTIONAL NEURAL
NETWORK**

yang disusun dan diajukan oleh

Hammam Afiq Murtadho

18.11.1931

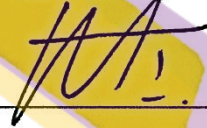
Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Desember 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Kusnawi, S.Kom., M. Eng.
NIK. 190302112



Windha Mega Pradnya D, M.Kom.
NIK. 190302185



Nafiatun Sholihah, S.Kom., M.Cs.
NIK. 190302524



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Desember 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hammam Afiq Murtadho
NIM : 18.11.1931

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Penyakit Paru-Paru Berdasarkan Citra X-Ray Menggunakan Metode Convolutional Neural Network

Dosen Pembimbing : Windha Mega Pradnya D, M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat **penyimpangan dan ketidakbenaran** dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Desember 2023

Yang Menyatakan,



METERAI
TEMPEL
3CF25AKX774961562

Hammam Afiq Murtadho

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan nikmat dan rahmat sehingga penulis dapat menyelesaikan skripsi dengan judul “Klasifikasi Penyakit Paru-Paru Berdasarkan Citra X-Ray Menggunakan Metode Convolutional Neural Network”. Skripsi ini disusun untuk memenuhi salah satu syarat mencapai derajat Sarjana Program Studi Informatika Universitas AMIKOM Yogyakarta.

Selama penyusunan skripsi ini penulis mendapat banyak dorongan, bantuan, serta doa dari berbagai pihak. Oleh karena itu penulis mengucapkan banyak terima kasih kepada:

1. Yth. Ibu Windha Mega Pradnya D, M.Kom., selaku Dosen Pembimbing sekaligus Dewan Penguji 2 yang telah memberikan kritik, saran serta arahan pada saat penyusunan skripsi ini.
2. Yth. Bapak Kusnawi, S.Kom., M. Eng. dan Ibu Nafiatun Sholihah, S.Kom., M.Cs., selaku Dewan Penguji 1 dan 3 yang telah memberikan kritik dan saran agar skripsi ini lebih baik.
3. Yth. Bapak Hanif Al Fatta, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Kedua orang tua penulis yang telah memberikan dukungan, nasehat serta doa selama hidup penulis, sehingga sampai saat ini skripsi penulis dapat terselesaikan dengan baik. Hanya kesuksesan dan kebaktian penulislah yang dapat membayarkan semua budi orang tua penulis, maka dengan skripsi ini penulis harap dapat memulai pembalasan budi terhadap orang tua.
5. Kakak penulis yang telah memberikan dukungan serta nasehat ketika pembuatan skripsi ini.

6. Semua teman-teman serta sahabat penulis yang sudah menemani penulis dari awal kuliah hingga saat ini dan selalu memberikan dukungan serta doa agar penulis tetap semangat dalam penyelesaian skripsi ini.
7. Seluruh dosen Jurusan S1 Informatika Universitas AMIKOM Yogyakarta yang telah memberikan ilmunya kepada penulis, sehingga penulis dapat memiliki basis pengetahuan untuk menyelesaikan skripsi ini.
8. Seluruh pihak lain yang telah membantu penulis dalam menyusun skripsi ini.

Selama penulisan skripsi ini penulis sadar masih banyak kekurangan dan kesalahan dikarenakan keterbatasan kemampuan dan sumber daya yang dimiliki oleh penulis. Oleh karena itu penulis mohon maaf atas kesalahan dan kekurangan dalam skripsi ini dan bersedia menerima segala kritik dan saran yang membangun agar lebih baik lagi ke depannya. Semoga skripsi ini dapat memberikan manfaat bagi siapa saja yang membaca.

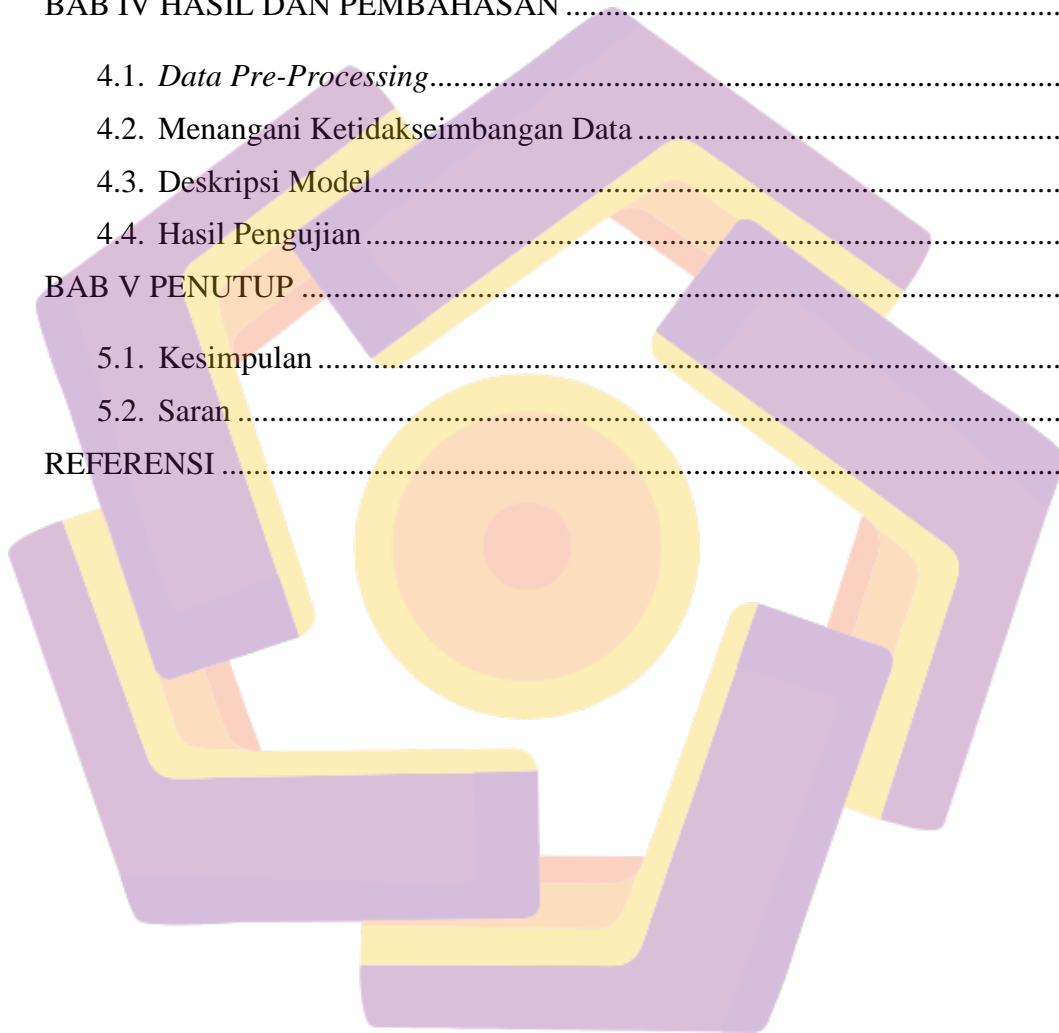
Yogyakarta, 18 Desember 2023

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	17
1.1. Latar Belakang	17
1.2. Rumusan Masalah.....	18
1.3. Batasan Masalah	18
1.4. Tujuan Penelitian	19
1.5. Manfaat Penelitian	19
1.6. Sistematika Penulisan	19
BAB II TINJAUAN PUSTAKA	21
2.1. Studi Literatur	21
2.2. Dasar Teori	27
2.2.1. Deep Learning.....	27
2.2.2. Citra X-Ray	28
2.2.3. <i>Convolutional Neural Network</i>	28
2.2.4. Model EfficientNetV2.....	32
2.2.5. Weighted Binary Cross-Entropy	32

2.2.6. Confusion Matrix	34
BAB III METODE PENELITIAN	37
3.1. Objek Penelitian.....	37
3.2. Alur Penelitian	37
3.3. Alat dan Bahan.....	40
BAB IV HASIL DAN PEMBAHASAN	42
4.1. <i>Data Pre-Processing</i>	42
4.2. Menangani Ketidakseimbangan Data	42
4.3. Deskripsi Model.....	45
4.4. Hasil Pengujian	46
BAB V PENUTUP	50
5.1. Kesimpulan	50
5.2. Saran	50
REFERENSI	51



DAFTAR TABEL

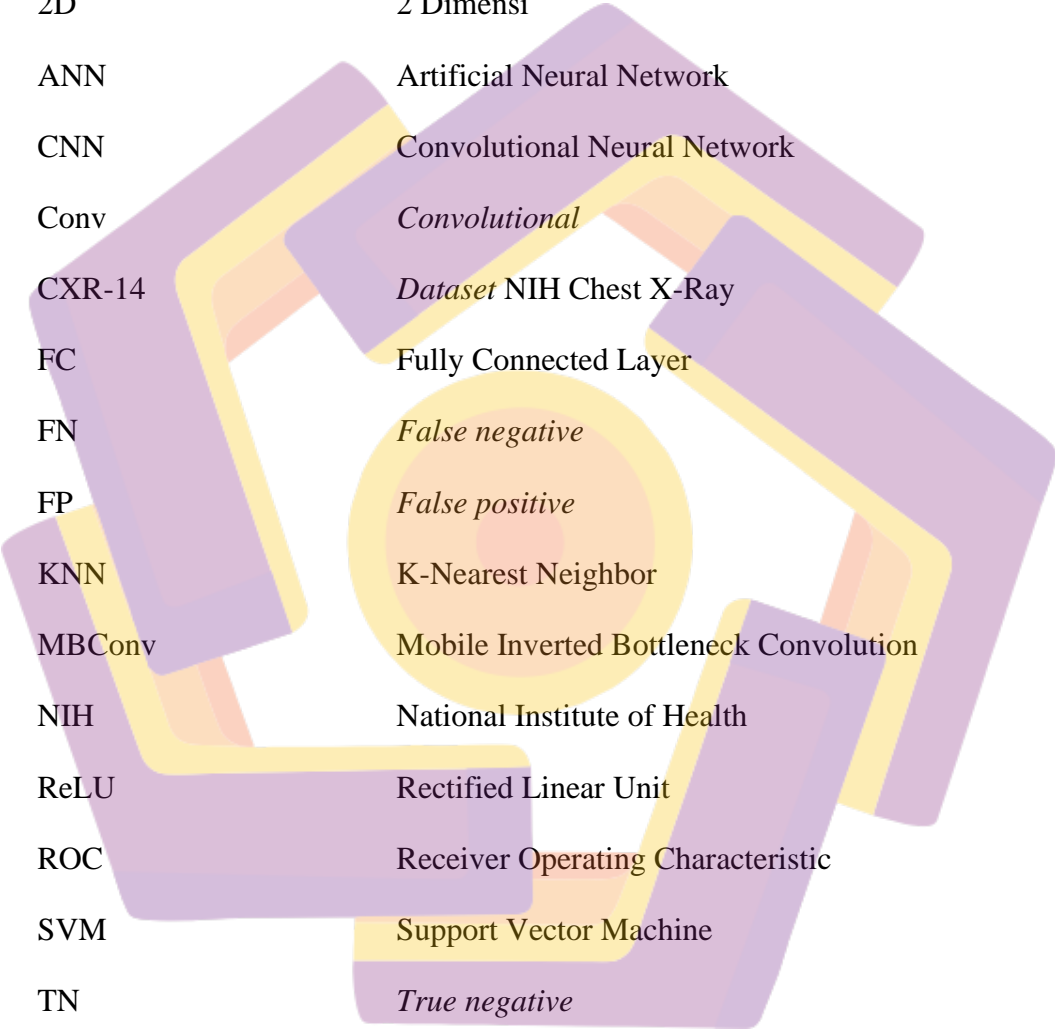
Tabel 2.1. Keaslian Penelitian	23
Tabel 2.2. Arsitektur EfficientNetV2-S	32
Tabel 3.1. Jumlah data pada tiap kelas	40
Tabel 4.1. Bobot positif masing masing kelas	44
Tabel 4.2. Arsitektur model	45
Tabel 4.3. Hasil pengujian sebelum <i>fine-tuning</i>	47
Tabel 4.4. Hasil pengujian setelah fine-tuning	48
Tabel 4.5. <i>Confusion Matrix Training</i>	48
Tabel 4.6. <i>Confusion Matrix Validation</i>	48
Tabel 4.7. Performa Model	49



DAFTAR GAMBAR

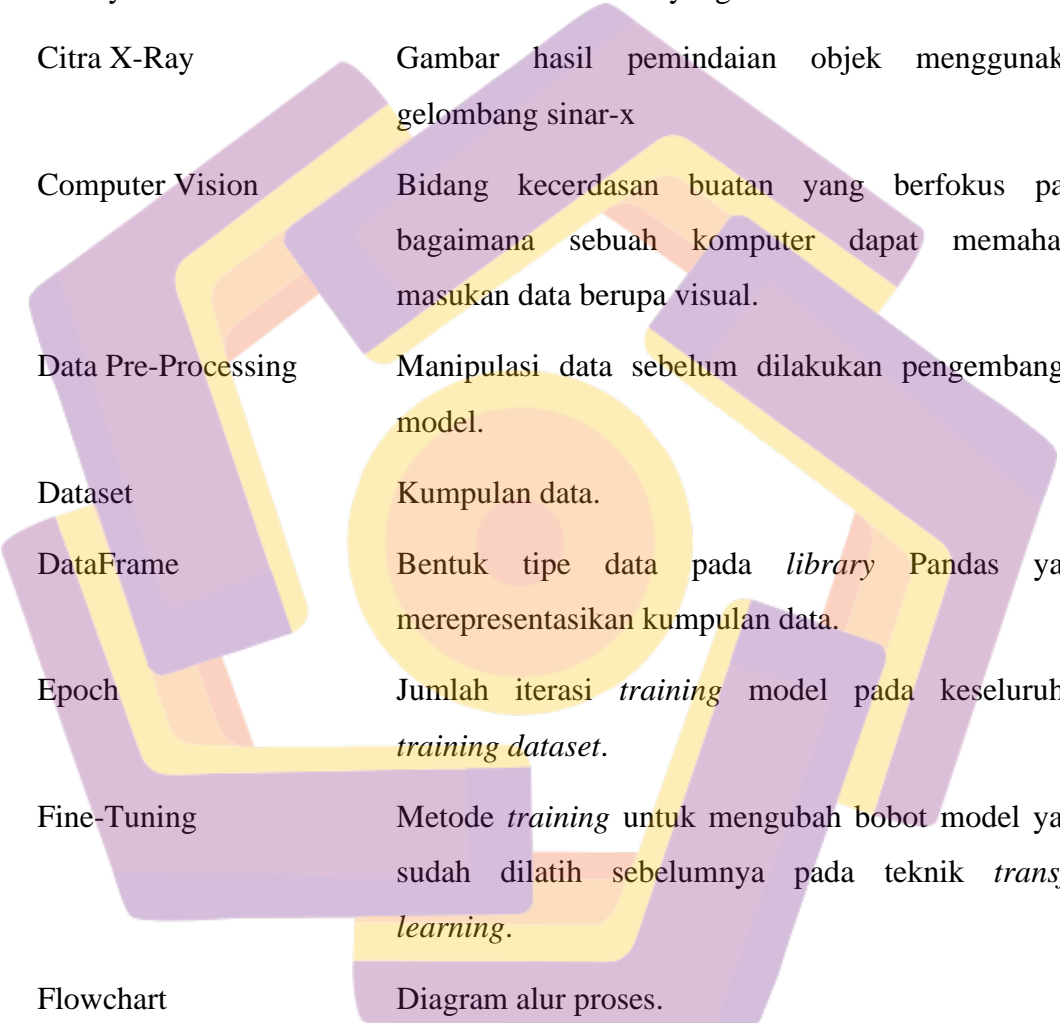
Gambar 2.1. Tahapan yang umum digunakan pada sebuah layer CNN	29
Gambar 2.2. Proses pada Convolutional Layer	29
Gambar 2.3. Jenis layer <i>pooling</i>	31
Gambar 2.4. Contoh Arsitektur CNN dengan <i>fully-connected layer</i>	31
Gambar 2.5. Contoh tabel <i>confusion matrix</i>	35
Gambar 3.1. Alur Penelitian	37
Gambar 4.1. Hasil augmentasi data	42
Gambar 4.2. Grafik persentase ... <i>dataset CXR-14 30%</i> .	43
Gambar 4.3. Grafik persentase ... <i>dataset CXR-14 100%</i> .	43
Gambar 4.4. Grafik akurasi sebelum <i>fine-tuning</i>	46
Gambar 4.5. Grafik <i>loss</i> sebelum <i>fine-tuning</i>	46
Gambar 4.6. Grafik akurasi setelah dilakukan <i>fine-tuning</i>	47
Gambar 4.7. Grafik <i>loss</i> setelah dilakukan <i>fine-tuning</i>	47
Gambar 4.8. Grafik ROC menunjukkan rata-rata skor 0.511	49

DAFTAR LAMBANG DAN SINGKATAN



\emptyset	Himpunan Kosong
Σ	Sigma
1D	1 Dimensi
2D	2 Dimensi
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
Conv	<i>Convolutional</i>
CXR-14	<i>Dataset</i> NIH Chest X-Ray
FC	Fully Connected Layer
FN	<i>False negative</i>
FP	<i>False positive</i>
KNN	K-Nearest Neighbor
MBCConv	Mobile Inverted Bottleneck Convolution
NIH	National Institute of Health
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
TN	<i>True negative</i>
TP	<i>True positive</i>

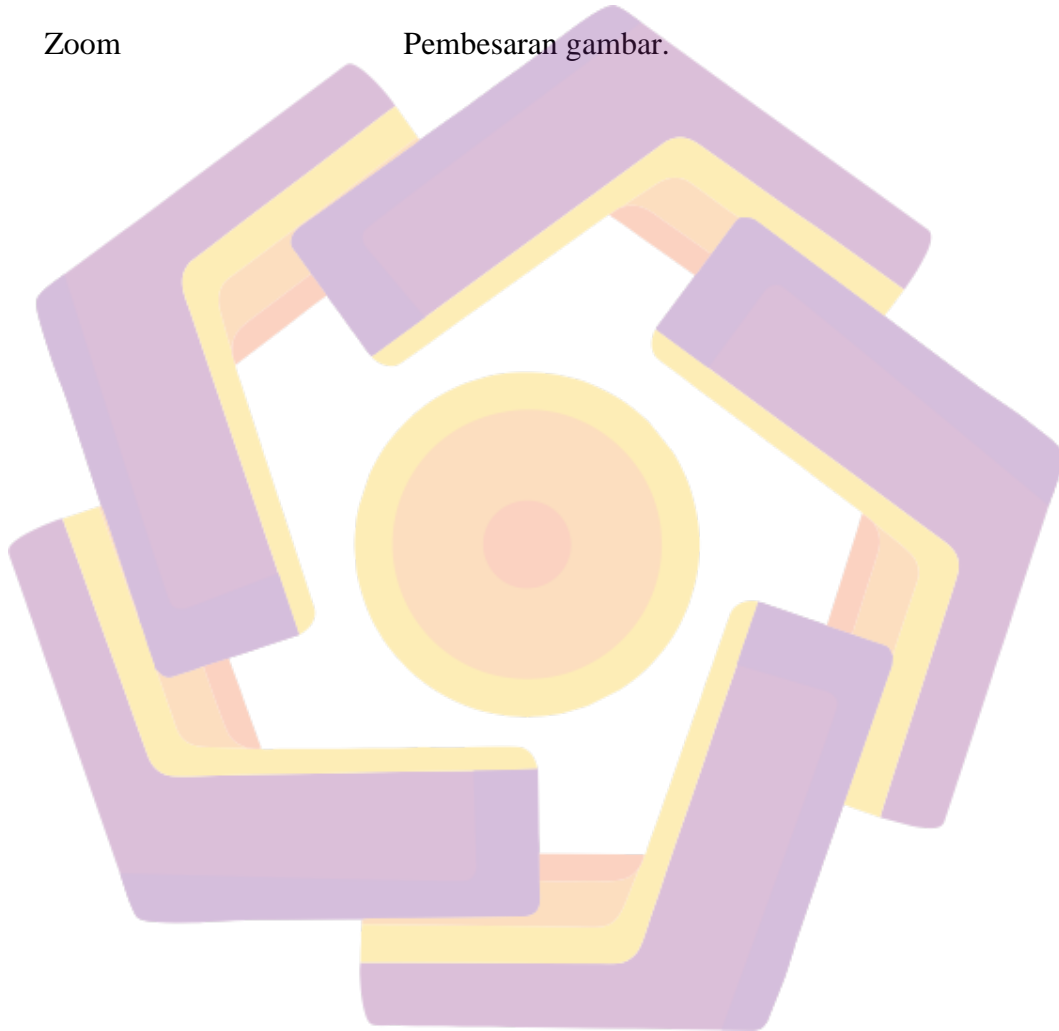
DAFTAR ISTILAH



Balancing	Proses penyeimbangan data.
Batch Size	Jumlah sampel data yang diproses oleh model dalam satu iterasi di dalam satu <i>epoch</i> .
Binary Classification	Jenis klasifikasi data yang memiliki 2 label eksklusif.
Citra X-Ray	Gambar hasil pemindaian objek menggunakan gelombang sinar-x
Computer Vision	Bidang kecerdasan buatan yang berfokus pada bagaimana sebuah komputer dapat memahami masukan data berupa visual.
Data Pre-Processing	Manipulasi data sebelum dilakukan pengembangan model.
Dataset	Kumpulan data.
DataFrame	Bentuk tipe data pada <i>library</i> Pandas yang merepresentasikan kumpulan data.
Epoch	Jumlah iterasi <i>training</i> model pada keseluruhan <i>training dataset</i> .
Fine-Tuning	Metode <i>training</i> untuk mengubah bobot model yang sudah dilatih sebelumnya pada teknik <i>transfer learning</i> .
Flowchart	Diagram alur proses.
Horizontal Flip	Membalikkan gambar pada sumbu horizontal.
Input	Masukan pada program.
Library	Kumpulan sumber daya tetap yang digunakan oleh perangkat lunak.

Loss	Fungsi yang digunakan untuk menghitung tingkat kesalahan yang dilakukan model dalam melakukan prediksi.
Machine Learning	Bidang studi kecerdasan buatan yang berfokus pada penggunaan data dan algoritma.
Model	Program <i>machine learning</i> .
Multi-Class Classification	Jenis klasifikasi data yang memiliki label eksklusif lebih dari dua.
Multi-Label Classification	Jenis klasifikasi data yang memiliki label non-eksklusif lebih dari satu.
Overfitting	Sifat model yang memiliki akurasi tinggi pada data <i>training</i> namun pada data baru akurasinya rendah.
Output	Keluaran program.
Random Guess	Teknik prediksi dengan hasil prediksi acak.
ROC Curve	Grafik yang menunjukkan performa model klasifikasi pada semua batasan klasifikasi.
Rotation	Rotasi gambar.
Scan X-Ray	Pemindaian objek menggunakan gelombang sinar-x.
Shear	Kemiringan sebuah gambar.
Testing	Tahapan pengetesan pada pengembangan model <i>machine learning</i> .
Training	Tahapan pelatihan pada pengembangan model <i>machine learning</i> .
Transfer Learning	Teknik <i>machine learning</i> yang menggunakan model yang sudah terlatih pada <i>dataset</i> lain.
Underfitting	Ketika model belum bisa mempelajari pola pada data dan tidak bisa memprediksi pada data baru.

Validation	Tahapan validasi model pada pengembangan model <i>machine learning</i> .
Vertical Flip	Membalikkan gambar pada sumbu vertikal.
X-Ray (Sinar-X)	Gelombang radiasi elektromagnetik yang berenergi tinggi.
Zoom	Pembesaran gambar.



INTISARI

Penyakit paru-paru adalah salah satu penyakit yang memiliki dampak serius terhadap kehidupan seseorang. Pasien yang diduga memiliki penyakit paru-paru dapat dilakukan *scan x-ray*, lalu didiagnosis oleh dokter apakah pasien benar memiliki penyakit paru-paru atau tidak. Tetapi tidak semua fasilitas kesehatan memiliki dokter yang mampu mendiagnosis penyakit paru-paru dengan reliabel. Oleh karena itu diperlukan sistem untuk membantu dokter untuk mendiagnosis penyakit paru-paru pada hasil *scan x-ray* dengan reliabel dan akurat. Banyak yang sudah mencoba membuat sistem deteksi penyakit paru-paru dengan berbagai macam teknik. Salah satunya adalah menggunakan metode jaringan saraf buatan di dalam bidang *computer vision*. Pada penelitian ini digunakan metode CNN dengan model EfficientNetV2 yang di latih menggunakan *dataset* dari NIH Chest X-Ray 14 (CXR-14). Dikarenakan *dataset* yang dipakai memiliki ketidakseimbangan data yang signifikan maka diperlukan metode *balancing*. Pada kalkulasi *loss* digunakan fungsi *weighted binary cross-entropy* untuk memberi bobot pada data minoritas. Hasil model setelah *training & fine-tuning* memiliki tingkat akurasi tertinggi 96,36%, tetapi memiliki nilai ROC rata-rata hanya 0,511. Skor ROC mendekati 0,5 menunjukkan model tidak lah lebih baik dari *random guess*.

Kata kunci: penyakit paru-paru, x-ray, cxr-14, EfficientNetV2, *weighted binary cross-entropy*.

ABSTRACT

Lung diseases is diseases that has serious impact on a person's life. Patients who are suspected of having lung disease can undergo an x-ray scan, then the doctor will diagnose whether the patient really has lung disease or not. But not all health-care facilities have doctors who can reliably diagnose lung diseases. Therefore, a system is needed to help doctors diagnose lung diseases on x-ray image reliably and accurately. Many have tried to create lung diseases detection system using various techniques. One of them is using artificial neural network methods in the field of computer vision. In this study, the CNN method was used with the EfficientNetV2 model which was trained using NIH Chest X-Ray 14 (CXR-14) dataset. Because the dataset has significant data imbalance, a balancing method is needed. In the loss calculation, the weighted binary cross-entropy function is used to give weight to minority data. The model results after training & fine-tuning have the highest accuracy of 96.36%, but have an average ROC value of only 0.511. An ROC score close to 0.5 indicates the model is no better than random guess.

Keyword: lung diseases, x-ray, cxr-14, EfficientNetV2, weighted binary cross-entropy.