

**PENERAPAN TEKNIK DATA AUGMENTATION UNTUK
PENGEMBANGAN MODEL CNN PADA KLASIFIKASI
CITRA X-RAY DALAM DETEKSI PENYAKIT PARU**
SKRIPSI

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



disusun oleh
ANANG MA'RUF
22.11.4769

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

**PENERAPAN TEKNIK DATA AUGMENTATION UNTUK
PENGEMBANGAN MODEL CNN PADA KLASIFIKASI
CITRA X-RAY DALAM DETEKSI PENYAKIT PARU**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
ANANG MA'RUF
22.11.4769

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN TEKNIK DATA AUGMENTATION UNTUK
PENGEMBANGAN MODEL CNN PADA KLASIFIKASI
CITRA X-RAY DALAM DETEKSI PENYAKIT PARU**

yang disusun dan diajukan oleh

ANANG MA'RUF

22.11.4769

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Februari 2026

Dosen Pembimbing,



Subektiningsih, S.Kom., M.Kom.
NIK. 190302413

HALAMAN PENGESAHAN

SKRIPSI

**PENERAPAN TEKNIK DATA AUGMENTATION UNTUK
PENGEMBANGAN MODEL CNN PADA KLASIFIKASI
CITRA X-RAY DALAM DETEKSI PENYAKIT PARU**

yang disusun dan diajukan oleh

ANANG MA'RUF

22.11.4769

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Emigawaty, M.Kom
NIK. 190302226

Dr. Hartatik, S.T., M.Cs.
NIK. 190302232

Subektiningsih, S.Kom., M.Kom.
NIK. 190302413



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusri, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Anang Ma'ruf
NIM : 22.11.4769

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Teknik Data Augmentation untuk Pengembangan Model CNN pada Klasifikasi Citra X-ray dalam Deteksi Penyakit Paru

Dosen Pembimbing : Subektiningsih, S.Kom., 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, 26 Februari 2026

Yang Menyatakan,

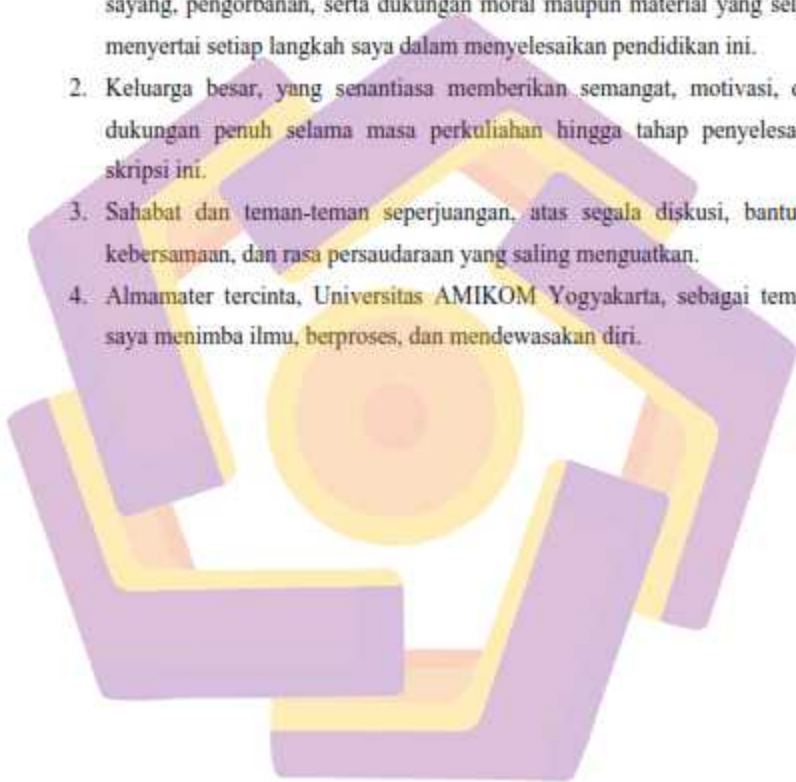


Anang Ma'ruf

HALAMAN PERSEMBAHAN

Dengan mengucapkan puji dan syukur kehadiran Allah SWT, karya tulis sederhana ini saya persembahkan sebagai wujud rasa hormat dan terima kasih yang mendalam kepada:

1. Kedua orang tua tercinta, atas segala doa yang tidak pernah putus, kasih sayang, pengorbanan, serta dukungan moral maupun material yang selalu menyertai setiap langkah saya dalam menyelesaikan pendidikan ini.
2. Keluarga besar, yang senantiasa memberikan semangat, motivasi, dan dukungan penuh selama masa perkuliahan hingga tahap penyelesaian skripsi ini.
3. Sahabat dan teman-teman seperjuangan, atas segala diskusi, bantuan, kebersamaan, dan rasa persaudaraan yang saling menguatkan.
4. Almamater tercinta, Universitas AMIKOM Yogyakarta, sebagai tempat saya menimba ilmu, berproses, dan mendewasakan diri.



KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan Skripsi ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat dalam mencapai derajat Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa penyelesaian skripsi ini tidak lepas dari bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Subektiningsih, M.Kom., selaku dosen pembimbing yang telah memberikan waktu, tenaga dan arahan selama penyusunan skripsi.
4. Seluruh Dosen dan Staf Akademik Universitas AMIKOM Yogyakarta yang telah memberikan bekal ilmu pengetahuan selama penulis menempuh pendidikan.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan karena keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan bagi pembaca pada umumnya.

Yogyakarta, 27 Februari 2026

Penulis

DAFTAR ISI

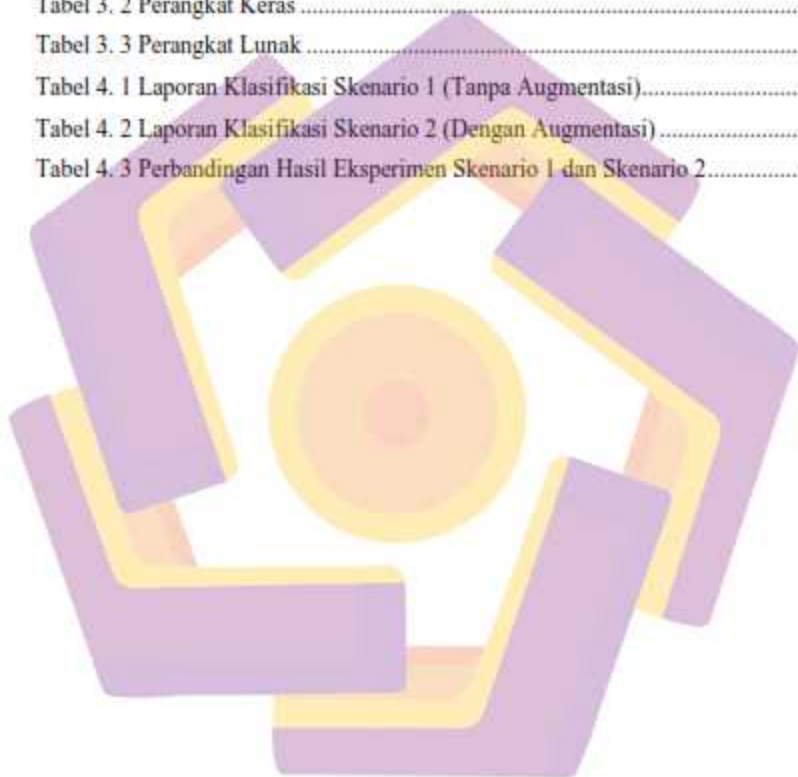
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	10

2.2.1	Penyakit Paru-paru	10
2.2.2	Citra Digital dan Sinar-X (<i>Chest X-Ray</i>)	11
2.2.3	Pengolahan Citra Digital (<i>Digital Image Processing</i>)	12
2.2.4	Deep Learning	14
2.2.5	Convolutional Neural Network (CNN)	16
2.2.6	Arsitektur MobileNetV2	16
2.2.7	Transfer Learning	18
2.2.8	Data Augmentation	20
2.2.9	Pengujian dan Evaluasi Model	22
BAB III METODE PENELITIAN		24
3.1	Objek Penelitian	24
3.2	Alur Penelitian	25
3.2.1	Pengumpulan Data	26
3.2.2	Pra-pemrosesan Data (<i>Preprocessing</i>)	26
3.2.3	Pembagian Data (<i>Data Splitting</i>)	26
3.2.4	Penerapan Teknik Augmentasi Data	26
3.2.5	Perancangan dan Pelatihan Model	27
3.2.6	Evaluasi dan Analisis	27
3.3	Alat dan Bahan	27
3.3.1	Perangkat Keras (<i>Hardware</i>)	27
3.3.2	Perangkat Lunak (<i>Software</i>)	28
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Pengumpulan data	30
4.2	Pra-Pemrosesan Data (<i>Data Preprocessing</i>)	32
4.2.1	Penyeragaman Ukuran Citra (<i>Image Resizing</i>)	32

4.2.2	Normalisasi Nilai Piksel (<i>Pixel Normalization</i>)	32
4.2.3	Pelabelan Kategori (<i>Label Encoding</i>)	33
4.3	Pembagian Data	33
4.4	Skenario Eksperimen	35
4.4.1	Skenario Tanpa Data Augmentation (<i>Baseline</i>)	35
4.4.2	Skenario Dengan Data Augmentation	36
4.5	Perancangan Model CNN	37
4.5.1	Arsitektur MobileNetV2 dan Transfer Learning	37
4.5.2	Modifikasi Lapisan Klasifikasi (<i>Custom Head</i>)	38
4.6	Pelatihan Model dan Optimizer Adam	39
4.6.1	Mekanisme Pelatihan Model	39
4.6.2	Fungsi dan Konfigurasi Optimizer Adam	40
4.6.3	Analisis Proses Konvergensi	41
4.7	Evaluasi Model	42
4.7.1	Confusion Matrix	42
4.7.2	Akurasi (<i>Accuracy</i>)	45
4.7.3	Loss (Categorical Crossentropy)	45
4.8	Analisis Komparatif	48
4.8.1	Analisis Perbedaan Akurasi	49
4.8.2	Stabilitas Model dan Indikasi Overfitting	50
BAB V PENUTUP		51
5.1	Kesimpulan	51
5.2	Saran	51
REFERENSI		53

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	7
Tabel 2. 2 Komponen Confusion Matrix	22
Tabel 3. 1 Distribusi Gambar.....	24
Tabel 3. 2 Perangkat Keras	27
Tabel 3. 3 Perangkat Lunak	28
Tabel 4. 1 Laporan Klasifikasi Skenario 1 (Tanpa Augmentasi).....	47
Tabel 4. 2 Laporan Klasifikasi Skenario 2 (Dengan Augmentasi)	47
Tabel 4. 3 Perbandingan Hasil Eksperimen Skenario 1 dan Skenario 2.....	49



DAFTAR GAMBAR

Gambar 2. 1 Langkah Pengolahan Citra Digital.....	13
Gambar 2. 2 Arsitektur Deep Learning.....	15
Gambar 2. 3 Arsitektur CNN.....	16
Gambar 2. 4 Arsitektur MobileNetV2.....	18
Gambar 2. 5 Alur proses Transfer Learning.....	19
Gambar 2. 6 Alur proses Data Augmentation.....	21
Gambar 3. 1 Alur Penelitian.....	25
Gambar 4. 1 Normal.....	31
Gambar 4. 2 Pneumonia.....	31
Gambar 4. 3 Tuberculosis (TB).....	31
Gambar 4. 4 Image Resizing.....	32
Gambar 4. 5 Pixel Normalization.....	33
Gambar 4. 6 Label Encoding.....	33
Gambar 4. 7 Pembagian Data.....	34
Gambar 4. 8 Data Augmentation.....	36
Gambar 4. 9 MobileNetV2.....	38
Gambar 4. 10 Custom Head.....	39
Gambar 4. 11 Pelatihan Model dan Optimizer Adam.....	40
Gambar 4. 12 Grafik Training Tanpa Augmentasi.....	41
Gambar 4. 13 Grafik Training Dengan Augmentasi.....	41
Gambar 4. 14 Confusion Matrix Tanpa Augmentasi.....	43
Gambar 4. 15 confusion Matrix Dengan Augmentasi.....	44
Gambar 4. 16 Perbandingan akurasi.....	45
Gambar 4. 17 Grafik Loss pada Model Tanpa Augmentasi.....	46
Gambar 4. 18 Grafik Loss pada Model dengan Augmentasi.....	46

DAFTAR LAMBANG DAN SINGKATAN



ANN	<i>Artificial Neural Network</i>
CAD	<i>Computer Aided Diagnosis</i>
CAP	<i>Community-Acquired Pneumonia</i>
CNN	<i>Convolutional Neural Network</i>
CXR	<i>Chest X-Ray</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
IDE	<i>Integrated Development Environment</i>
JPEG	<i>Joint Photographic Experts Group</i>
PNG	<i>Portable Network Graphics</i>
ReLU	<i>Rectified Linear Unit</i>
TB	<i>Tuberkulosis</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
WHO	<i>World Health Organization</i>

DAFTAR ISTILAH



<i>Batch Size</i>	Jumlah sampel citra yang diproses oleh algoritma secara bersamaan sebelum model memperbarui parameter bobotnya.
<i>Batch Size</i>	Jumlah sampel citra yang diproses oleh algoritma secara bersamaan sebelum model memperbarui parameter bobotnya.
<i>Confusion Matrix</i>	Tabel matriks evaluasi yang membandingkan label yang dihasilkan oleh model dengan label kebenaran dasar (<i>ground truth</i>) dari dataset.
<i>Data Augmentation</i>	Teknik manipulasi data (seperti rotasi, zoom, <i>flip</i>) untuk memperluas variasi dataset latih secara sintesis guna mencegah <i>overfitting</i> .
<i>Deep Learning</i>	Cabang dari <i>Machine Learning</i> yang menggunakan jaringan saraf tiruan berlapis ganda (<i>hidden layers</i>) untuk mengekstraksi dan mempelajari representasi data kompleks secara otomatis.
<i>Epoch</i>	Satu siklus pelatihan komplit di mana seluruh dataset latih telah diproses dan dilewatkan melalui model algoritma.
<i>Fine-tuning</i>	Proses penyesuaian bobot pada lapisan model <i>pre-trained</i> agar lebih relevan dan spesifik dengan dataset baru yang sedang diteliti.
<i>Learning Rate</i>	Parameter laju pembelajaran yang mengontrol seberapa agresif model menyesuaikan bobotnya setiap kali menghitung kesalahan.

One-Hot Encoding

Teknik konversi yang mengubah label kategori teks (seperti Normal, TB) menjadi format vektor angka biner yang dapat dikomputasi oleh algoritma.

Overfitting

Kondisi malfungsi model di mana akurasi saat proses pelatihan sangat tinggi, namun gagal/akurasinya rendah saat memprediksi data uji yang baru.

Transfer Learning

Teknik mendaur ulang model yang sudah dilatih pada dataset raksasa (seperti ImageNet) untuk menyelesaikan masalah baru pada dataset yang ukurannya lebih kecil.



INTISARI

Penyakit paru-paru seperti Tuberkulosis (TB) dan Pneumonia memiliki tingkat mortalitas yang tinggi, sehingga memerlukan deteksi dini yang cepat dan akurat melalui pemeriksaan X-ray, yang mana secara manual rentan terhadap subjektivitas dan kelelahan visual radiolog. Meskipun teknologi *Deep Learning*, khususnya *Convolutional Neural Network (CNN)*, mampu mengekstraksi fitur visual kompleks secara otomatis, penerapannya sering terkendala oleh terbatasnya dataset medis yang memicu terjadinya *overfitting*. Oleh karena itu, penelitian ini bertujuan untuk menerapkan teknik *Data Augmentation* dalam pengembangan model CNN menggunakan arsitektur MobileNetV2 untuk mengklasifikasikan citra X-ray ke dalam tiga kelas: Normal, Pneumonia, dan Tuberkulosis. Penelitian ini menggunakan 2.210 dataset citra sekunder yang dibagi menjadi 80% data latih dan 20% data validasi, dengan menerapkan skenario augmentasi geometri dinamis yang meliputi rotasi, *zoom*, *shear*, dan *flip* selama proses pelatihan. Evaluasi performa yang diukur menggunakan *Confusion Matrix*, Akurasi, Presisi, *Recall*, dan *F1-Score* menunjukkan bahwa pengembangan model dengan augmentasi mencapai akurasi validasi sebesar 95,93%, dibandingkan model *baseline* (tanpa augmentasi) sebesar 97,28%. Meskipun secara angka sedikit lebih rendah, penerapan teknik *data augmentation* terbukti efektif bertindak sebagai mekanisme regularisasi yang memberikan tingkat stabilitas model yang lebih baik dan secara konsisten mampu mencegah *overfitting* pada dataset yang terbatas.

Kata kunci: *Deep Learning, Convolutional Neural Network, MobileNetV2, Data Augmentation, Klasifikasi X-Ray Paru.*

ABSTRACT

Lung diseases such as Tuberculosis (TB) and Pneumonia have high mortality rates, necessitating fast and accurate early detection through X-ray examinations, which manually are prone to the subjectivity and visual fatigue of radiologists. Although Deep Learning technology, particularly Convolutional Neural Networks (CNN), can extract complex visual features, its application is often constrained by limited medical datasets, which triggers overfitting. Therefore, this study aims to implement Data Augmentation techniques in the development of a CNN model using the MobileNetV2 architecture to classify X-ray images into three classes: Normal, Pneumonia, and Tuberculosis. This study utilized 2,210 secondary image datasets divided into 80% training data and 20% validation data, applying dynamic geometric augmentations including rotation, zoom, shear, and flip during the training process. Performance evaluation measured using Confusion Matrix, Accuracy, Precision, Recall, and F1-Score showed that the developed model with augmentation achieved a validation accuracy of 95.93%, compared to the baseline model (without augmentation) at 97.28%. Although slightly lower numerically, the application of data augmentation techniques proved effective as a regularization mechanism, demonstrating better model stability and consistently preventing overfitting on limited datasets.

Keyword: *Deep Learning, Convolutional Neural Network, MobileNetV2, Data Augmentation, Lung X-Ray Classification.*